(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 September 2002 (26.09.2002)

PCT

(10) International Publication Number
WO 02/074961 A1

(51) International Patent Classification[7]: C12N 15/12, 15/00, 1/21, C12P 21/02, G01N 33/48

(21) International Application Number: PCT/US02/05109

(22) International Filing Date: 14 March 2002 (14.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/810,173 15 March 2001 (15.03.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/810,173 (CIP)
Filed on 15 March 2001 (15.03.2001)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ZHOU, Ping** [US/US]; 7595 Newcastle Drive, Cupertino, CA 95014 (US). **GOODRICH, Ryle** [US/US]; 4896 Sandy Lane, San Jose, CA 95124 (US). **ASUNDI, Vinod** [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). **ZHANG, Jie** [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). **ZHAO, Qing, A.** [CN/US]; 1556 Kooser Road, San Jose, CA 95118 (US). **REN, Feiyan** [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). **XUE, Aidong, J.** [CN/US]; 1621 South Mary Avenue, Sunnyvale, CA 94087 (US). **YANG, Yonghong** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **MA, Yunqing** [CN/US]; 280 W. California Avenue, #206, Sunnyvale, CA 94086 (US). **YAMAZAKI, Victoria** [JP/US]; 883 Portwalk Place, Redwood Shores, CA 94065 (US). **CHEN, Rui-Hong** [US/US]; 1031 Flying Fish Street, Foster City, CA 94404 (US). **WANG, Zhiwei** [CN/US]; 836 Alturas Avenue, #B36, Sunnyvale, CA 94085 (US). **GHOSH, Malabika** [IN/US]; 620 Iris Avenue, Apartment 133, Sunnyvale, CA 94086 (US). **WEHRMAN, Tom** [US/US]; CCSR Mol Pharm 3210, 269 W. Campus Drive, Stanford, CA 94305 (US). **WANG, Jian-Rui** [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). **WANG, Dunrui** [CN/US]; 12252 Pepper Tree Lane, Poway, CA 92064 (US). **DRMANAC, Radoje, T.** [US/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

WO 02/074961 A1

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. BACKGROUND OF THE INVENTION

1.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

1.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

2. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1 – 526. The polypeptide sequences are designated SEQ ID NOS: 527 - 1052. The nucleic acids and polypeptides are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1 – 526 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 527 - 1052. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1 – 526 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1 – 526. The sequence

information can be a segment of any one of SEQ ID NO: 1 – 526 that uniquely identifies or represents the sequence information of SEQ ID NO: 1 – 526.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-526 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-526 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1–526; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1–526; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1–526. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under

stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1–526; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-526; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These

techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form

the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products.

Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

3. DETAILED DESCRIPTION OF THE INVENTION

3.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules.

The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of

oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-526.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-526. The sequence information can be a segment of any one of SEQ ID NOs: 1-526 that uniquely identifies

or represents the sequence information of that sequence of SEQ ID NO: 1-526. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol)

and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions,

deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of

glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted"

proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of

this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

3.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1 – 526; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 527 - 1052; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1 - 526. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1 – 526; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing as SEQ ID NO: 527 - 1052; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 527 - 1052. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1 – 526 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1 - 526 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1 - 526 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1 - 526, or complements thereof, which fragment is

greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1 - 526, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1 - 526 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1 - 526, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to

create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-526, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 526 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 526 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein

recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus*

subtilis, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

3.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1 - 526, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that

comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 527 -1052 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1 - 526 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1 - 526, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil,

5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic

acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

3.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1 - 526). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a

DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

3.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous

promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result

in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

3.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 527 -1052 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1 - 526 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NOs: 1 – 526 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO 527 -1052 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 527 -1052 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 527 - 1052.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing

primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography,

ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 527 -1052.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.*, as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art

(see, *e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

3.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

3.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for

example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

3.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the

present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein

produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

3.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the

regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

3.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and

truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

3.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify

polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

3.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

3.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine

Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

3.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-

pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell

lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and

cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

3.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

3.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the

invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

3.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may

also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue

transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal

models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a

T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte

Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

3.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be

useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

3.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell

population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

3.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al.,

Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

3.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin

keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these

individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

3.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

3.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays.

Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the

"hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

3.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.,* increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2)

natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

3.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease,

inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

3.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

3.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

3.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or

circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

3.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which

appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

3.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of

the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

3.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

3.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

3.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti- inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hyl, IL-1 Hy2,

anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or

other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

3.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery; the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

3.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an

isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture

with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active

ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent.

Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist

in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering

the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer

matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

3.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et

al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

3.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

3.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 527 -1052, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for

targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the

invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

3.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

3.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are

sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can

be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

3.13.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are

those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

3.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in

culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

3.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

3.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have

different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example,

bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H

domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

3.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

3.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

3.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as

dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

3.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NOs: 1 - 526 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NOs: 1 - 526 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present

invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic

acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

3.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

3.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention

under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or

antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

3.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

3.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a

polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NOs: 1 - 526, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of

Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

3.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1 - 526. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NOs: 1 - 526 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled

nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

3.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8)

3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-$MeIm_7$), is then added to a final concentration of 10 mM 1-$MeIm_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-$MeIm_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing

solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

3.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76

clones showed that *CviJI*** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

3.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell

plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

4.0 EXAMPLES

4.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In

some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

4.2 EXAMPLE 2

Novel Nucleic Acids

The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 119, gb pri 119, and UniGene version 119) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 121, gb pri 121, UniGene version 121, Genpept release 121) and the amino acid version of Genseq released February 15, 2001. Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1- 526.

Table 1 shows the various tissue sources of SEQ ID NO: 1-526.

The nearest neighbor results for polypeptides encoded by SEQ ID NO: 1-526 (i.e. SEQ ID NO: 527 - 1052) were obtained by a BLASTP (version 2.0al 19MP-WashU) search against Genpept, Geneseq and SwissProt databases using BLAST algorithm. The nearest neighbor result showed the closest homologue with functional annotation for SEQ ID NO: 527 - 1052. The translated amino acid sequences for which the nucleic acid

sequence encodes are shown in the Sequence Listing. The homologues with identifiable functions for SEQ ID NO: 527 - 1052 are shown in Table 2 below.Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), polypeptides encoded by SEQ ID NO: 1-526 (i.e. SEQ ID NO: 527 - 1052) were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) polypeptides encoded by SEQ ID NO: 1-526 (i.e. SEQ ID NO: 527 - 1052) were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the product of all the e-value of similar domains found, the pFam score for the identified domain within the sequence, number of similar domains found, and the position of the domain in the SEQ ID NO: being interrogated.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ ID NO: 1-526 (i.e. SEQ ID NO: 527 - 1052). Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by Altschul et al, (Nucl. Acids. Res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA,) which is an automated sequence and structure searching procedure (http://www.msi.com/), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows, "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as annotated by the PDB files (http:/www.rcsb.org/PDB/); start and end amino acid position

of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David Eisenberg's laboratory (US patent no. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:12502-13597. The verify score produced by GeneAtlas normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

Verify score (normalized) = (raw score – 1/2 high score)/(1/2 high score)

The PMF score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potential (MFP). As given in Table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good model. Similarly, a PMF score between 0 to 1.0, with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may also be determined by one of skill in the art based on all the information in Table 5 taken in totality.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determined from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al., as reference, were obtained for the polypeptide sequences. Table 6 shows the position of the last amino acid of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

Table 7 correlates each of SEQ ID NO: 1-526 to a specific chromosomal location.

Table 8 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-526, novel polypeptide sequences SEQ ID NO: 527 - 1052, and their corresponding priority nucleotide sequences in the priority application USSN 09/810,173, herein incorporated by reference in its entirety.

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
adipocytes	Stratagene	ADP001	39 49 68 84 103-104 117 124 186 188-189 221 247 272 307 312 336-337 353 356 369 434 461 495 509
adrenal gland	Clontech	ADR002	11 14 25 30 39 83 90 92 100 108 111 131 133 137 144 148 155 164 170-173 184 196 206 244-245 254 260 266 273 301 317 330 349 359 383 392 397-398 401 411-414 423 442 466 468 486 510-511 518
adult brain	BioChain	ABR012	47 262
adult brain	BioChain	ABR013	60 205
adult brain	Clontech	ABR001	17 39 55 61 95 124 137 153 186 233 247 252 287-288 307 322 353 377 380 388 412-414 448 482 505-506 511
adult brain	Clontech	ABR006	9 17 26 32 38 41 61 77 81 83 87 95 106 117 134 137 143 153-154 158 163 175-176 179 181 193 205 217 224-227 235 254 257 262 277 308 340 342 359 369 376 389-391 419 433 442 446-447 458-461 466 474 482 484 497-498 509 512 515
adult brain	Clontech	ABR008	2 4 7 12 17-18 24-25 28-29 32 35-38 44 46-48 50 57 62-63 65-68 70 74-75 77 84 96 101 103-104 107-109 112-113 117 120 125 127 144 151-153 158 163 166-167 170-175 178 181-182 185 187 191 193 196 200-201 204 209-210 223 225 231 239-243 247-248 257 259 262 264-266 276-277 279-280 282 289-290 311-312 321-322 326 331 337-338 342 346-347 349 353 356 358 360 366 369 375 380 389-391 405 408 411-414 426-427 442 449 452-454 456 458 463 473-476 480 482 489 493 495 498 503 505-506 510-512 515 521
adult brain	Clontech	ABR011	394
adult brain	GIBCO	AB3001	9 13 21 32 34 49 58 61 77 92 98 124 138-141 154 205 248 254 282 289 298 309 323 326 342 371 412-414 461 475
adult brain	GIBCO	ABD003	9 15-16 18 24 26 32 34 39 54 60-61 66 68 79 96 98 109 112 117 120 124 131 140 143-144 153-154 162 170-173 181 195-196 201 205 223 231 233-234 252 257 273 287-288 298 300 313 317 323 326 345-346 369 371 376 379 383-384 386 397 405 411-414 418 442 495 497 501 511 521
adult brain	Invitrogen	ABR014	65 125 184 247 307 338 467 490 509 513
adult brain	Invitrogen	ABR015	12 34 60 73 127 140 287 417 445
adult brain	Invitrogen	ABR016	3 24 34 136 177 248 307 452 474
adult brain	Invitrogen	ABT004	29 39 47 65-66 83 87 97 107 143 151-152 156 163 166-167 193 196 217 220-221 254 266 281 307 317 334 378 382 389 397 412-414 430 473 509
adult heart	GIBCO	AHR001	5-6 11 15-16 18-20 23 34 39 41 48 50 62-63 65 70 77 84 86 92 95-100 103-104 107 109 111 114 118 124-125 127 142-144 154 162 165-167 170-175 178 181-182 186 188 191 193-197 200 206-207 217 221 224 228 247 257 266 273-275 281 287-288 317 337 340 346 353 355 362-363 369 374 376-377 382 384-385 390-391 397-398 400 411-414 423 434 440 474 482 489 498 500-502 509-510 513
adult kidney	GIBCO	AKD001	5-6 11-12 14-16 19 22 24 27 32 34 39 41 46-47 49 51 53 55 58 62-63 68 77 80 83-84 91-92 98 100-107 110 116 119 125-127 137 144-147 154 160 162 165 178 181-182 188-189 193 207 210 215-217 231-233 240 247-249 254 257 264 273-274 287-288 298 306 321 323 326 330 334 340 342 346 353 367 371 376 382 384-385 394 397 400 411-414 429-430 444 446-447 456 461 467 474-475 482 489 495-498 509-511 514 516 524-525
adult kidney	Invitrogen	AKT002	1 18 27 34 58 66 77 101 107 124 129 131 136-137 146 155 181-182 196 206 217 264 266 274-275 288 291 320 326 334 375-376 394 400-401 408 411-414 418 423 435-437 444 452 458 473 481-482 501 504 509 519
adult liver	Clontech	ALV003	32 74-75 94 137 247 420 516

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
adult liver	Invitrogen	ALV002	6 12 18 23 25 34 49 65 74-75 80 87 94 98 118 122 133 137 151-152 163 170-173 186 197 223 236 246-247 254 258 266 285 326 344 353 370 383 387 399 412-414 452 456 460 462 466 473 475 497 519
adult lung	GIBCO	ALG001	15-16 18 27 34 47 65 72 74-75 83 92 127 137 155 185 210-212 215-216 248 288 318 326 331 337 382 400 434 461 474 492 495 516
adult lung	Invitrogen	LGT002	5 11-12 14 16 18-19 24 26 30 32 34-36 39 46 49 55 57 66-67 73-75 80 84 92 97-99 103-105 108 112 120 124-125 134 150 166-167 169-173 179-180 182 186 188 193 196 202-203 210 212 215-217 221 225 231 246-247 254 256 266 273 281-282 288 307 309 317-318 326 331 338 342 346 348 353 356 365-366 375-376 381 385 389 397-398 411-414 418 426 434 452 456 475 489-490 495 501 503-504 508-509 521
adult spleen	Clontech	SPLc01	17 22 25 54 71 108 117 121 130 133 153 184 207 226-227 254 257 281-282 331 346 364 384 398 406 416 461 512
adult spleen	GIBCO	ASP001	15-16 22 24 26 34 41 77 96 103-104 107 111-112 121 124 142 144 155 158 163 182 206-207 215-216 255 281 287 326 337 342 364 370 398 411-414 434 456 473-474 495 511
bladder	Invitrogen	BLD001	35-36 77 103-104 124 144 218 281 287 337 367 369 376 430 434 460 509
bone marrow	Clonetech	BMD007	32
bone marrow	Clontech	BMD001	2 5 9 12 15 17-18 20 24-25 27 30 34 38 54-58 68-72 77 88-91 95 103-104 110 112 122 124 155 162 165 176 178 181-182 186 188 193 199 204 215-217 221 230 233 246 254 274 288 292 305 307 309 326 331 340 342 349 364 376 379 389-391 401 411-414 416 441 446-448 489 497-498 500 503 513-514 516 518 524
bone marrow	Clontech	BMD004	346 460
bone marrow	GF	BMD002	4 17-18 23-25 27-28 30 32 35-36 38 47 51 53 57 71 74-75 77 87 90-92 95 103-104 107-108 113 117 122-125 133 137 148-149 151-152 154-155 170-173 178 181-182 184 186 189 191 196 198 209 215-216 221 231 233 250 254 266 272 276 281 283 287 301 317 326 330-331 337 342 346 349 356 364-366 371 379 392 394 396 402 406 408 411-414 421-422 433 435-438 442 461 467-468 475 489 495 498 501 503 505-506 509-510 512 514 517-518
cervix	BioChain	CVX001	5-6 18 20 24 30 32 42 44 55-56 66 68 72 84 92 96 99-100 110-111 120 131 134 137 144 146 151-152 162 165 170-173 175-176 181-182 184 186 190 193 195 197 207 210 214-216 238 246-247 254 266 272-273 275 282 287 291-293 298 317 321 323 326 333 340 342 353 355 365 367 369-370 378 382 411-414 418 423 434 438 452 456 458 460-465 473-474 476 479 492 498 500 504 507 510 524
colon	Invitrogen	CLN001	11 13 34 81 100 105 126 184 186 196 254 317 328 330 349 400 412-414 426 460 466 510 525
diaphragm	BioChain	DIA002	226-227
endothelial cells	Strategene	EDT001	2 13-14 16-19 22 24 26-27 30-31 34-36 47 49 53 58 62-63 65-68 73 80 83 85-86 92 96 98 100 102 106-108 114 117-118 125-126 132 137 142 144 148-149 164 166-167 170-173 175 178 181-182 188-190 196-197 206 213-214 217 221 231 246-247 254 257 266 273 288 306-307 309 313 318 323 326 334 337 340 342 355 366 369 371 375-376 379-382 389 400 406 409 411-414 423 426 429 431 440 445 452 456 461 467-468 474 482 490 503-504 508-510 514 516
fetal brain	Clontech	FBR001	39 87 247 353 375 452 460 513
fetal brain	Clontech	FBR004	181 205 393
fetal brain	Clontech	FBR006	1 7-8 12 17-19 24 27 29-30 32 34-36 46-49 53 58 62-63 70 77-78 85 95-96 103-104 107-108 120 125 127 134 151-153 164 166-167 175-176 182 184-185 189 196 201 204 217 223 225 229 231 242 245 247

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			253-254 264 266 269-270 275 280-281 287 294-295 304-305 321 326 329 331 346 353 355-356 359 369 375 379 381 389-391 394 411 418 423-427 430 440 442 445 449-450 452 454 456 461 463 469 474 478 481-482 493 495 502 504-506 511-512 518
fetal brain	GIBCO	HFB001	5 11-12 15-16 18 20 24 27-28 30 34-36 46 53 58-67 69 84 92 97 100 112 114 120 124 128 134-136 138 143 151-154 159-167 182 184 186 188-190 193 196 205 207 217 221 223 233 248 264 266 273-274 282 285 287 305 307-308 326 338 340 342 349 366-367 371 375 379 389-391 397 400 411-414 431 442 452 467 476 480 482 489 492 497-498 503 508-509 511
fetal brain	Invitrogen	FBT002	13 15 18-19 25 37 42 46 60 65-66 74-75 118 132 137 140 150-153 175 185 196 203 222-223 235 247-248 266 298 307 331 353 366 382 397 430 452 481-482 495 508-509 511
fetal heart	Invitrogen	FHR001	6 15 18-19 24 26 29 37 46 57-58 74-75 77-78 81 96 103-104 114 127 134-135 151-153 160 164 178 181 184 186 191 201 204-205 207 224 242 245 247 253-254 257 273 276 281-282 287-288 309 312 317 326 338 353 356 363 370 376-377 382 390-391 394 400 406 408 411-414 427-430 439-440 453 474 478 489-490 495 498 501 510-512 515 525
fetal kidney	Clontech	FKD001	17 39 92 97 99 133 193 203-205 318 326 371 397 401 411-414 448
fetal kidney	Clontech	FKD002	27-28 46 48-49 53 69-70 81 94 105 117 131 137 181-182 196 200 205 221 226-227 247 254 258 329 337-338 373 381 397 415 431 451-452 463 488 503 511-512 515
fetal liver	Clontech	FLV002	19 170-173 223 298 401
fetal liver	Clontech	FLV004	4 19 25-26 29 32 37-38 46 53 80-81 92 96 100-101 103-104 108 114 124 127 136 153 178 181 184-185 199 208 215-216 257 272 287 298 306 309 326 376 396 401 442 446 453 461 467 474 497 510 512
fetal liver	Invitrogen	FLV001	12 16 25 32 44 60 77 80 117 137 144 188 230 246-247 266 272 281 298 342 353 382 401 412-414 449 460 482 495 519
fetal liver-spleen	Soares	FLS001	2-21 23-43 45-55 58 65 67 69-70 72-81 83 85-86 92-94 96-97 100 103-108 110 115 120 124-125 131 133 137 144 146 149-155 158-159 165 175 178 180-182 185-186 189 191-193 196 210 215-216 228-230 238 246-248 254 264 266 272-273 282-283 285 288 292 298 305 307 309 317-318 321 323 326 330 334-337 339-341 345-346 351 353 355 359 365-366 370-371 375-376 382 384-386 389 395-402 411-414 426 434 438 441-442 444 449 458 467 474-475 481-482 489-490 492 495 497 501 503-512 514 516 519 522 525
fetal liver-spleen	Soares	FLS002	2-3 5-6 9 11-12 15-16 18-20 23-28 31 35-36 38-41 47-49 51-55 57-60 65 68 73-75 77 80 83 90 93 97-98 100-101 107-108 114 120 124 127-128 131 133 137 143-144 148 150-152 155 157 163 166-167 174 177 179 181-182 184 187-191 196 200 215-216 226-227 229-231 241 246-248 254 258 266 272 285 287-288 307 312 316 326 335-342 346 348 350-356 366 370-371 376 379 382 386 389 398 401 405 409 411-414 434 441-445 448-449 452 458 460 466 471 474-475 481-482 489-490 497 501 516 518 521
fetal liver-spleen	Soares	FLS003	6 16 21 48 65 72 84 98 110 114 124 208 215-216 229 254 286 288 307 317 336-337 356 366 370 397 401 405-408 434 444-447 455 493 497-498 501 504-506
fetal lung	Clontech	FLG001	65 137 237 247 281 312 334 434 510
fetal lung	Invitrogen	FLG003	49 66 77 105 121 182 246-248 281 294 302 337 353 366 401 412-414 460
fetal muscle	Invitrogen	FMS001	9 23 53 84 95 118 281 322-323 331 336 346 355 366 401 446-447 461 473 498 509 519
fetal muscle	Invitrogen	FMS002	23 25 28-29 48 58 92 103-104 124 127 131-132 201 217 247 255 257

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			276 281 316 323 326 337 353 373 411 429 431 446-447 453-454 474 490 498 502 512 519
fetal skin	Invitrogen	FSK001	5 9 15-16 18 24 26 28 30 32 35-36 40 48 62-63 66 77 87 95-96 98 103-104 107-109 120 124 131-132 141 170-173 175 177 182 186 198 204 226-227 235 251 266 273 281 285 287 295 302 309 313 320-328 332-333 336 346 349 353 355 366-367 369 375 385-386 389-391 397 401 411 434 442 452 456 460 467 501 509-510 512
fetal skin	Invitrogen	FSK002	4 24-26 31 46 48 53 68 71 74-75 77 81 87 109-110 117 151-152 170-173 178 181-182 185-186 193 196 204 209 215-216 225-227 245 247 253 275-276 287 307 326 328 331 333 337 353 369 373 379 390-391 412-414 418 432 439-440 452-454 463 467 475 489 495-496 502-503 505-506 510 512 515
fibroblast epilepsy	Julio_m	EPM001	357
fibroblast epilepsy	Julio_m	EPM004	357
fibroblasts	Julio_m	BAC001	484
infant brain	NULL	IBM002	13 42 48 61 77 170-173 184 190 308 444 456 467
infant brain	NULL	IBS001	26 60 84 100 137 143 170-173 175 184 281 315 366 376 397 489 507
infant brain	Soares	IB2002	9 13 16 18 20 22 24 26 30-31 34 37-38 45 47-48 54 60-63 66 69 77 80-81 83-84 95-96 99 103-104 111 117 119 121 124-125 127 139-140 154-155 160 162-163 168 175-176 179 182 184-185 196 200-201 205 218 220 226-227 247 252 259-260 266 273 281 287-288 307-308 317 326 331 337 340 342 346 349 353 365 369-371 375 383-384 390-391 397-398 426 434 442 444 446-447 456 458 460-461 467 473-474 481 489 492 495 497-498 501 505-507 509-511 525
infant brain	Soares	IB2003	2 13-14 17 24-25 30 38 49 61 66 77 87 95 107 130 137 140 143-144 153-154 163 175-176 184-185 196 200-201 205 207 223 245 247-248 254 259 266 273 281 287-288 307-308 317-318 326 331 338 341 346 353 371 383-384 397 411 442 456 458 460 489 492 495 497 501 507 510 512 515
leukocytes	Clontech	LUC003	5 77 112 137 165 181-182 272 307 376 416 453 508-509 512
leukocytes	GIBCO	LUC001	5 13-15 18-20 24-25 27 32 34 37 39 43 46-47 53 55-56 58 64 67-68 70 74-77 84 87 96 101 103-104 108-115 123-126 131 135 137 143-144 150 153 155 164-167 169 178-179 181-182 184 188-190 196 200 207 210 212 215-216 221 223 235 248 254 257 267 274 281-283 287 302 306-307 309 312 316-317 321 326 331 337 340 342 349 364-366 371 375-376 379 382 389-391 394 396-397 405-406 411-414 416 426 429 434 442 444 452 457-458 464-465 467 470 489 495 501 503-506 509 511-513 524
lung	Strategene	LFB001	6 11 13 15 41 46 56 84 92 112 143 154 178 181 190 197 202 217 282 307 312 336 365 389 456 474 482 484 504
lymph node	Clontech	ALN001	18 71 122 155 176-177 202 326 338 411
lymphocyte	ATCC	LPC001	5 15 24-25 29 39 44 53-55 70-71 87 92 96 107 112 117 120 125 131 137 144 155 165 181-182 210 217 254 266-267 272 283 288 317 321 342-343 346 365 370 375 379 384 394 396 411 442 448 453 461 467-468 474 478 493 496 501 503-504 513
macrophage	Invitrogen	HMP001	24 69 113 129 137 144 287 326 389 396 398 406 467 510
mammary gland	Invitrogen	MMG001	15-18 24-26 30 32 35-37 39 44 49 62-63 65-66 72 77-78 83 87 100-101 103-105 107 109 112 114 117 131-132 137 144 146 151-153 157-158 170-173 182 187-188 190 196-197 223 234-235 240 243 246-248 254 266 272 281 283 287 298 300 302 317 319 326 330-333 337 341-342 353 355-356 371 375 380-382 385 397 400 411-414 423 434 442

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			445 452 456-457 460-461 465 473 475 495 501 507-510 516 519 521 525
melanoma	Clontech	MEL004	18 39 50 73 92 118 124 127 208 212 247 285 303-304 317 322 326 342 353 452 473-474 492
*Mixture of 16 tissues - mRNA	Various Vendors	CGd010	60 77 94 322 338 473 478-479 496 519
*Mixture of 16 tissues - mRNA	Various Vendors	CGd011	39 77 243 247 352 401 412-414 471 480 500
*Mixture of 16 tissues - mRNA	Various Vendors	CGd012	13 18 20 25-26 30 39 46 50 56 59 65 72 77 80-81 95 99 108 110 124 144 148 189 194 215-216 225 232 241 243 247 284 287 299 326 331 337 351-352 368 380 390-391 401 412-414 418 460 467 471 493 499-503 516
*Mixture of 16 tissues - mRNA	Various Vendors	CGd013	26 58 81 105 127 284 331
*Mixture of 16 tissues - mRNA	Various Vendors	CGd015	4 18 34 39 60 67 71 106 147 180 207 254 331 367 370-371 401 456 497 501 503 507-509 512
*Mixture of 16 tissues - mRNA	Various Vendors	CGd016	2 29-30 77 112 131 143 175 184 248 259 307 335 359 397 401 409 505-506 524
neuron	Strategene	NTD001	3 8 11-13 45 69 77 79 81 131 137 139-140 166-167 179 207 295 307 317 361 366 423 444 497 514 520
neuron	Strategene	NTR001	77 81 95 103-104 111 163 181-182 342 353 375 379 446-447 456 460 467 495
neuronal cells	Strategene	NTU001	17 39 79 95 111 117 140 151-152 182 266 305-306 358 369 373 375 398 430 448 458 467 475 509 514
ovary	Invitrogen	AOV001	2 5-6 8-9 12 15-16 18 20 24-25 27-28 30 34 39 44 48 54 58 61 65 67-69 74-75 77 84 86-87 95 97-98 101 103-105 107 110-112 114 118 120 125-127 131 134 137-138 142 144 148-150 153 155-156 162 164-165 169-173 175-187 189-190 193 197 199-200 205 207 210 215-219 221 225-228 231 246-247 254 264 266 272 274-275 281-282 288 298 307 309 313 317-318 321 323 326 331 336-338 340-342 346 349 353 355-356 365-366 369-370 373-376 378 380-382 389 394 411-414 418 423 434 442 444 452 455-456 458 467-468 473-474 477 481 489 492 496-497 500 504 507 509-510 515-516 521 524-525
pituitary gland	Clontech	PIT004	12 14 137 151-152 164 189 266 380 461 467 513 516 521
placenta	Clontech	PLA003	24 71 84 92 96 103-104 178 182 184 246 262 289 304 317 326 331 333 337 385 431 433 440 452 493 503 511-512
placenta	Invitrogen	APL001	151-152 182 215-216 247 340
placenta	Invitrogen	APL002	24 34 49 80 83 107 112 125 153 190 247 353 397 400 510
prostate	Clontech	PRT001	15 28 53 80 96 105 112 124-125 141 181 184 196 246-248 281 298 353 368 382 474 499 524
rectum	Invitrogen	REC001	18 78 80 83 105 196 226-227 248 266 275 281 296-297 321 366 369 390-391 397-398 406 411-414 460 489 509-510
saliva gland	Clontech	SALS03	482
salivary gland	Clontech	SAL001	25 39 41 124 202 268 299 338 340 353 355 365 381 411 418 430 489-490 498 501 516
skeletal muscle	Clontech	SKM001	11 23 182 186 217 226-227 247 353 378 386 411 498 513 525
skin fibroblast	ATCC	SFB001	16
small intestine	Clontech	SIN001	12 18 20 24 26 30 35-36 39 48 53 62-63 74-75 86 92 99-100 105 107-

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			108 112-114 120 125 137 142 153-154 165 169-173 175 182 185-186 204 206-207 210 215-216 221 229 231 246-247 249 251 254-255 266 272 281-282 285 287 298 307 316 326 330 337 340 342 349 351 353 356 367 369 371 376 382 385 389 394 396 400 403-404 410 412-415 417 456-457 460-461 466 482 484 489 501 512-513 516 525
spinal cord	Clontech	SPC001	22 34 41 51 66 88 121 124 133 137 155 158 178 181-182 196 214-216 229-230 247-248 250 254 261 269-271 281 318 353 367 369 371 416 444 458 461 496 501 508 510-512 520
stomach	Clontech	STO001	9 16 55 86 165 251 254 274 282 323 355 385 410 434 457 482 501 507
testis	GIBCO	ATS001	13-15 17-18 24 41 46 66 77 80 107-108 110 120 131 154 162 178 185 233 246 272 281 287-288 306 317 342 365 394 400 411 418 427 434 444 489 495 504 509
thalamus	Clontech	THA002	32 39 60 68 126 137 144 154 185 190 247 252 254 273 308 321 341 349 353 371 397 400 430 466 475 521
thymus	Clonetech	THM001	14 17 25 28 30 34 39 49 53-54 61 76 87 100 124 128 137 151-152 158 182 196 202 215-216 246-247 254 261 274 281 298 316 322-323 340 346 349 353 364 366 369-371 376 384 389 408 411-414 438 444 455 467 489 501 504 509 516 524
thymus	Clontech	THMc02	4 18 25 27 34-36 38 46-47 53-54 64 71 74-75 77 81 87-88 92 96 108 137 155 170-173 180 184 196 200 202 211 225-227 229 231 233 239 254 262 272 281 283-284 287 310 316 333 337 356 366 369 373 375-376 390-391 397 406 411 431 442 459-460 467 473-474 482 501 503 509 512 516 518 520 524
thyroid gland	Clontech	THR001	5 9 11-12 14 16-19 24-25 27 29-30 34 42 46-48 55 57-58 61 67 69 77 88 92 96 100 114 120 124 128-129 131 133-134 137 151-155 165 170-173 175 177 182 196 206 215-216 231 247 249 251 253-255 263-264 266 272-275 282 285 288 307 309 330-331 337 340 345 349 353 365-367 369 371-372 376 381 396-397 409-414 433-434 440 444 452 456 467 475 497 500 509 511 513 515 524-525
trachea	Clontech	TRC001	18 24 70 126 174 215-216 238 251 286 365 383 456 489 510 520 524
umbilical cord	BioChain	FUC001	9 15 17-18 22 26 29-30 34 39 41 47 58 70 72 96 99 103-104 112 114 120 124 128-129 151-152 157 161 170-173 182 186 207 215-216 228 238 246-247 254 273 285 287 300 302 307 314 317 321 326 329-333 336-338 342 353 367 369 378-379 382 389-391 401 406 444 448 452 461 465 468 474 489 492 508-509 512 521 524
uterus	Clontech	UTR001	47 84 111 114 197 211 246-247 273 281 307 353 384 412-414 442 489 504
young liver	GIBCO	ALV001	15-16 23 38 67 92 96 101 114 120 130 137 154 165 176 182 184 186 209 254 337 340 366-367 382 405 411-414 429 452 474 497

*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human so\spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
527	gi9798452	Homo sapiens	mRNA for putative capacitative calcium channel (trp7 gene).	4470	100
527	gi5326854	Mus musculus	receptor-activated calcium channel	4392	98
527	gi2295903	Homo sapiens	Human putative calcium influx channel (htrp3) mRNA, complete cds.	3529	81
528	AAG89238	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 358.	545	100
528	AAG93320	Homo sapiens	NISC- Human protein HP10515.	545	100
528	gi13620915	Homo sapiens	bMRP63 mRNA for mitochondrial ribosomal protein bMRP63, complete cds.	545	100
529	AAW78211	Homo sapiens	HUMA- Human secreted protein encoded by gene 86 clone HTWCT03.	333	88
529	gi7294596	alt 2	CG4300 gene product [Drosophila	65	31
529	gi7294595	alt 1	CG4300 gene product [Drosophila	65	31
530	AAB95369	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17686.	2361	100
530	gi10435142	Homo sapiens	cDNA FLJ13215 fis, clone NT2RP4001447.	2361	100
530	gi16041164	Macaca fascicularis	hypothetical protein	1576	89
531	gi13625172	Homo sapiens	5-HT receptor mRNA, complete cds.	1615	93
531	gi10503978	Homo sapiens	clone SP329 unknown mRNA.	1615	100
531	gi7300419	Drosophila melanogaster	CG17796 gene product	96	27
532	gi10438219	Homo sapiens	cDNA: FLJ21986 fis, clone HEP06248.	1425	99
532	AAO13496	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27388.	1125	99
532	ABB11720	Homo sapiens	HYSE- Human novel protein, SEQ ID NO:2090.	725	97
533	gi4929685	Homo sapiens	CGI-108 protein mRNA, complete cds.	269	98
533	gi12838900	Mus musculus	putative	269	98
533	AAY65253	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1414.	265	96
534	gi220500	Mus musculus	NDPP-1 protein	65	29
534	gi6679028	Mus musculus] > [Mus musculus	NPC derived proline rich protein 1	65	29
535	AAG02210	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6291.	397	98
536	gi7573295	Homo sapiens	Human DNA sequence from clone RP1-238O23 on chromosome 6. Contains part of the gene for a novel protein similar to PIGR (polymeric immunoglobulin receptor), part of the gene for a novel protein similar to rat SAC (soluble adenylyl cyclase), ESTs, STSs and GSS, complete sequence.	389	75
536	gi4140400	Rattus norvegicus	soluble adenylyl cyclase	176	47
536	AAB81929	Homo sapiens	STRD Human soluble adenylyl cyclase.	172	45

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
537	AAY10830	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	246	100
537	gi13815145	Sulfolobus solfataricus	Hypothetical protein	68	37
537	gi15898682	Sulfolobus solfataricus] > [Sulfolobus solfataricus	Hypothetical protein	68	37
538	gi12841269	Mus musculus	putative	503	84
538	AAY13186	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 200.	406	97
538	AAW67825	Homo sapiens	HUMA- Human secreted protein encoded by gene 19 clone HELBW38.	369	100
539	AAS15817_aa 1	Homo sapiens	SAAT/ Human cDNA encoding prostate specific protein SSH9.	730	94
539	AAU10191	Homo sapiens	SAAT/ Human prostate specific protein SSH9.	730	94
539	AAB58298	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 636.	730	94
540	AAB43589	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1034.	913	100
540	gi5817181	Homo sapiens	mRNA; cDNA DKFZp566E104 (from clone DKFZp566E104); partial cds.	745	99
540	gi7512814	Homo sapiens	hypothetical protein DKFZp566E104.1 - human (fragment) >	745	99
541	AAB58235	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 573.	1480	100
541	gi5410296	Homo sapiens	homeobox prox 1 mRNA, complete cds.	1267	100
541	gi4929667	Homo sapiens	CGI-99 protein mRNA, complete cds.	1267	100
542	gi7108913	Homo sapiens	glucocorticoid receptor AF-1 coactivator-1 mRNA, partial cds.	1818	100
542	AAM66710	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27016.	513	66
542	AAM54312	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26417.	513	66
543	AAA61620_aa 1	Homo sapiens	MITO- cDNA encoding human ubiquitin-conjugating enzyme rapUBC.	275	100
543	AAZ10849_aa 1	Homo sapiens	DAND TIA-1 binding protein 1 (TIABP1) gene.	275	100
543	AAV51398_aa 1	Homo sapiens	DAND Human TIABP1 genomic DNA.	275	100
544	AAB43887	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1332.	1183	100
544	gi533111	Canis familiaris	signal peptidase complex 25 kDa subunit	1130	95
544	gi12856773	Mus musculus	putative	1129	95
545	gi6841242	Homo sapiens	HSPC296	567	99
545	gi12842164	Mus musculus	putative	564	97
545	gi7293870	Drosophila melanogaster	CG6884 gene product	236	45

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
546	gi3043652	Homo sapiens	mRNA for KIAA0564 protein, partial cds.	5065	100
546	gi3875726	Caenorhabditis elegans	similar to nir like gene involved in denitrification~cDNA EST yk12a1.3 comes from this gene~cDNA EST yk7e7.3 comes from this gene~cDNA EST yk7e7.5 comes from this gene~cDNA EST yk34c7.3 comes from this gene~cDNA EST yk12a1.5 comes from this gene~cDNA EST yk24f12.5 comes from this gene~cDNA EST yk34c7.5 comes from this gene~cDNA EST yk154e5.3 comes from this gene~cDNA EST yk212d10.3 comes from this gene~cDNA EST yk212d10.5 comes from this gene~cDNA EST yk225b7.3 comes from this gene~cDNA EST yk225b7.5 comes from this gene~cDNA EST yk243b7.5 comes from this gene~cDNA EST yk349d4.5 comes from this gene~cDNA EST yk367e8.3 comes from this gene~cDNA EST yk367e8.5 comes from this gene~cDNA EST yk420f3.3 comes from this gene~cDNA EST yk420f3.5 comes from this gene~cDNA EST yk529f9.5 comes from this gene~cDNA EST yk565d10.5 comes from this gene	1447	34
546	gi10728542	Drosophila melanogaster	c12.2 gene product	1005	56
547	gi12052936	Homo sapiens	mRNA; cDNA DKFZp566E2324 (from clone DKFZp566E2324); complete cds.	955	100
547	gi10439692	Homo sapiens	cDNA: FLJ23112 fis, clone LNG07874.	580	100
547	gi6692513	Hepatitis B virus	large S protein	81	32
548	AAY07902	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 51.	322	88
548	gi4008342	Caenorhabditis elegans	predicted using Genefinder~contains similarity to Pfam domain: PF01496 (V-type ATPase 116kDa subunit family), Score=925.6, E-value=4.6e-275, N=1~cDNA EST yk15f10.3 comes from this gene~cDNA EST yk15f10.5 comes from this gene~cDNA EST yk224h11.3 comes from this gene~cDNA EST yk223d1.3 comes from this gene~cDNA EST yk287c7.3 comes from this gene~cDNA EST yk321h11.3 comes from this gene~cDNA EST yk224h11.5	66	39

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			comes from this gene~cDNA EST yk223d1.5 comes from this gene~cDNA EST yk287c7.5 comes from this gene~cDNA EST yk321h11.5 comes from this gene		
548	gi7496564	Unknown	hypothetical protein C26H9A.1 - Caenorhabditis elegans >	66	39
549	gi17389834	Homo sapiens	Similar to RIKEN cDNA 2310035L15 gene, clone MGC:23953 IMAGE:4292862, mRNA, complete cds.	1024	100
549	gi12844552	Mus musculus	putative	906	89
549	AAM93823	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3881.	727	46
550	AAH26493_aa 1	Homo sapiens	BOST- Human low density lipoprotein binding protein 1 (LBP-1) gene.	697	94
550	AAH26492_aa 1	Homo sapiens	BOST- Human low density lipoprotein binding protein 1 (LBP-1) cDNA.	697	94
550	AAB82802	Homo sapiens	BOST- Human low density lipoprotein binding protein 1 (LBP-1).	697	94
551	gi12698332	Homo sapiens	C/EBP-induced protein mRNA, complete cds.	2084	100
551	gi14150747	Mus musculus	GIG18	641	43
551	gi5739567	Homo sapiens	BAC clone RP11-505D17 from 7p22-p21, complete sequence.	635	44
552	gi11761611	Homo sapiens	kinesin-like protein RBKIN1 (RBKIN) mRNA, complete cds, alternatively spliced.	6087	99
552	gi11761613	Homo sapiens	kinesin-like protein RBKIN2 (RBKIN) mRNA, complete cds, alternatively spliced.	5852	96
552	gi12054030	Homo sapiens	mRNA for KINESIN-13A1 (KIN13A gene).	5771	95
553	gi17391063	Homo sapiens	Similar to RIKEN cDNA 1500032H18 gene, clone MGC:21379 IMAGE:4509694, mRNA, complete cds.	1311	100
553	gi12837824	Mus musculus	putative	1083	83
553	gi7292416	Drosophila melanogaster	CG14985 gene product	383	35
554	gi12857727	Mus musculus	putative	1260	94
554	gi6851256	Mus musculus	protein tyrosine phosphatase-like protein PTPLB	1242	93
554	AAB59515	Homo sapiens	HUMA- Human secreted protein BLAST search protein SEQ ID NO: 104.	1092	100
555	AAM93439	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3078.	1266	100
555	gi16741367	Homo sapiens	clone MGC:17276 IMAGE:4180160, mRNA, complete cds.	1266	100
555	gi15079907	Homo sapiens	Similar to secretory carrier membrane protein 4, clone MGC:19661 IMAGE:3161979, mRNA, complete	1266	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			cds.		
556	gi16507984	Human endogenous retrovirus K115	putative env	430	48
556	gi4185944	Human endogenous retrovirus K	env protein	429	47
556	gi3150438	Human endogenous retrovirus K	pol-env	429	47
557	AAB98212	Homo sapiens	NANF- Human early endosome antigen 1 isomer (hEEA1-iso) SEQ ID NO:7.	1129	100
557	gi9963835	Homo sapiens	AD024 mRNA, complete cds.	1129	100
557	gi12834062	Mus musculus	putative	717	78
558	gi12847029	Mus musculus	putative	1082	76
558	AAY60569	Homo sapiens	META- Human normal bladder tissue EST encoded protein 241.	1073	100
558	gi12854670	Mus musculus	putative	525	80
559	gi15824269	Homo sapiens	NEDD4-like ubiquitin ligase 3	64	34
559	gi2662159	Homo sapiens	KIAA0439	64	34
560	AAB43895	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1340.	814	100
560	gi5231141	Homo sapiens	sin3 associated polypeptide (SAP18) mRNA, complete cds.	804	100
560	gi2108210	Homo sapiens	sin3 associated polypeptide p18 (SAP18) mRNA, complete cds.	804	100
561	gi17061811	Homo sapiens	C21orf57 isoform A protein (C21orf57) mRNA, partial cds, alternatively spliced.	1102	80
561	AAM25823	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1338.	938	97
561	gi17061813	Homo sapiens	C21orf57 isoform B protein (C21orf57) mRNA, partial cds, alternatively spliced.	804	64
562	gi17061811	Homo sapiens	C21orf57 isoform A protein (C21orf57) mRNA, partial cds, alternatively spliced.	818	75
562	AAM25823	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1338.	687	97
562	AAY48371	Homo sapiens	META- Human prostate cancer-associated protein 68.	674	96
563	AAB93239	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12243.	1630	100
563	gi15928956	Homo sapiens	clone MGC:22951 IMAGE:4872309, mRNA, complete cds.	1630	100
563	gi14042582	Homo sapiens	cDNA FLJ14798 fis, clone NT2RP4001313, weakly similar to MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM40.	1630	100
564	AAB94479	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15153.	1521	100
564	gi10434979	Homo sapiens	cDNA FLJ13111 fis, clone	1521	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NT2RP3002566.		
564	gi14043295	Homo sapiens	clone IMAGE:3534358, mRNA, partial cds.	1448	100
565	gi15620831	Homo sapiens	mRNA for KIAA1886 protein, partial cds.	1420	99
565	gi13276647	Homo sapiens	mRNA; cDNA DKFZp761I2123 (from clone DKFZp761I2123); complete cds.	1420	99
565	AAY86184	Homo sapiens	HELI- Nuclear transport protein clone hfb2007 protein sequence.	1364	99
566	gi4321787	Mus musculus	6-pyruvoyl-tetrahydropterin synthase	156	42
566	gi12832727	Mus musculus	putative	156	42
566	gi202561	Rattus norvegicus	6-pyruvoyl-tetrahydropterin synthase	148	41
567	gi13477179	Homo sapiens	hypothetical protein FLJ10342, clone MGC:12937 IMAGE:2820292, mRNA, complete cds.	1036	100
567	gi12804363	Homo sapiens	hypothetical protein FLJ10342, clone MGC:4366 IMAGE:2822886, mRNA, complete cds.	1036	100
567	gi12653941	Homo sapiens	hypothetical protein FLJ10342, clone MGC:2740 IMAGE:2822886, mRNA, complete cds.	1036	100
568	gi9280047	Macaca fascicularis	unname dprotein product	596	97
568	gi14532556	Arabidopsis thaliana	AT5g57360/MSF19_2	91	33
568	gi13487068	Arabidopsis thaliana	Adagio 1	91	33
569	AAY87333	Homo sapiens	INCY- Human signal peptide containing protein HSPP-110 SEQ ID NO:110.	543	93
569	AAY12883	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:473.	226	86
569	AAY12868	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:458.	168	81
570	gi17389322	Homo sapiens	Similar to NICE-5 protein, clone MGC:21212 IMAGE:3907760, mRNA, complete cds.	130	65
570	AAY73387	Homo sapiens	INCY- HTRM clone 3340290 protein sequence.	122	75
570	AAG73684	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4448.	76	45
571	gi9280156	Macaca fascicularis	unnamed protein product	168	82
571	AAO11992	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 25884.	76	50
571	AAO08245	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22137.	70	43
572	gi12666208	Homo sapiens	Human DNA sequence from clone RP11-103J18 on chromosome 13 Contains ESTs, STSs, GSSs and a CpG island. Contains two novel genes and the 3' part of a novel gene similar to	490	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			mouse MO25, complete sequence.		
572	AAU09964	Homo sapiens	MILL- Human cytidine deaminase-like protein from clone 26934.	425	100
572	AAG04055	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8136.	425	100
573	gi12654927	Homo sapiens	clone MGC:5509 IMAGE:3453623, mRNA, complete cds.	1201	100
573	gi13905264	Mus musculus	Similar to hypothetical protein MGC5509	1034	85
573	gi9022437	Xenopus laevis	ashwin	241	41
574	gi13477177	Homo sapiens	Similar to RIKEN cDNA 1500032A17 gene, clone MGC:12936 IMAGE:2820022, mRNA, complete cds.	1128	100
574	gi12851027	Mus musculus	putative	1012	89
574	AAG04038	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8119.	506	92
575	AAG93293	Homo sapiens	NISC- Human protein HP10659.	1343	100
575	gi15929856	Homo sapiens	Similar to RIKEN cDNA 0610011N22 gene, clone MGC:21397 IMAGE:3852440, mRNA, complete cds.	1343	100
575	gi13097141	Mus musculus	RIKEN cDNA 0610011N22 gene	1156	82
576	AAO07956	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 21848.	74	38
576	gi5917666	Zea mays	extensin-like protein	74	40
576	gi3980411	Arabidopsis thaliana	putative proline-rich protein	74	39
577	AAG89212	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 332.	324	100
577	gi4980816	Thermotoga maritima	hypothetical protein	72	36
577	gi9294037	Arabidopsis thaliana		67	45
578	gi13324963	Caenorhabditis elegans	Hypothetical protein F37B4.9	73	41
578	gi6677927	Mus musculus	sphingosine phosphate lyase 1	65	30
579	gi12856429	Mus musculus	putative	869	66
579	gi16549784	Homo sapiens	cDNA FLJ30562 fis, clone BRAWH2004731.	763	99
579	gi12848379	Mus musculus	putative	659	62
580	AAY94526	Homo sapiens	INCY- Human lysine-rich statherin protein.	342	96
580	gi438731	Mesomys hispidus	cytochrome b	75	39
580	gi1478112	Sciurus aberti	cytochrome b	73	38
581	AAY73460	Homo sapiens	GEMY Human secreted protein clone yk14_1 protein sequence SEQ ID NO:142.	416	100
582	AAY07790	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 47.	294	100
582	gi7107077	Porcine	envelope glycoprotein	63	55

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		reproductive and respiratory syndrome virus			
582	gi15231798	Arabidopsis thaliana	putative protein	63	34
583	gi6331397	Homo sapiens	mRNA for KIAA1287 protein, partial cds.	6081	99
583	gi12053113	Homo sapiens	mRNA; cDNA DKFZp434H1220 (from clone DKFZp434H1220); complete cds.	6081	99
583	gi12850252	Mus musculus	putative	1511	93
584	gi13623583	Homo sapiens	clone IMAGE:3939163, mRNA, partial cds.	610	99
584	AAG01516	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5597.	522	98
584	gi12654201	Homo sapiens	clone IMAGE:3449838, mRNA, partial cds.	458	100
585	gi16519031	Homo sapiens	putative tetracycline transporter-like protein mRNA, complete cds.	535	99
585	gi2506078	Mus musculus	tetracycline transporter-like protein	535	99
585	gi12836216	Mus musculus	putative	535	99
586	gi16550027	Homo sapiens	cDNA FLJ30760 fis, clone FEBRA2000536, weakly similar to Homo sapiens paraneoplastic cancer-testis-brain antigen (MA5) mRNA.	2043	100
586	gi14043275	Homo sapiens	clone MGC:15827 IMAGE:3507248, mRNA, complete cds.	2043	100
586	AAB12529	Homo sapiens	SLOK Human Ma5 protein SEQ ID NO:13.	754	46
587	gi9929997	Macaca fascicularis	hypothetical protein	856	93
587	AAB45027	Homo sapiens	HUMA- Human secreted protein encoded by gene 3.	76	52
587	gi13359187	Homo sapiens	mRNA for KIAA1657 protein, partial cds.	73	44
588	gi13559239	Homo sapiens	Human DNA sequence from clone RP5-842G6 on chromosome 20. Contains the 3' end of a novel gene, the 3' end of the gene for a novel protein similar to SEL1L (sel-1 (suppressor of lin-12, C.elegans)-like), ESTs, STSs and GSSs, complete sequence.	815	100
588	AAY38477	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 23.	712	75
588	gi16769652	Drosophila melanogaster	LD45826p	618	54
589	gi9971051	Homo sapiens	Human DNA sequence from clone RP11-526K24 on chromosome 20. Contains a novel gene, the 5' end of a novel gene, two CpG islands, ESTs, GSSs and STSs, complete sequence.	585	100
589	AAG01028	Homo sapiens	GEST Human secreted protein, SEQ ID	579	96

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 5109.		
589	gi6782267	Caenorhabditis elegans	cDNA EST yk536g11.3 comes from this gene~cDNA EST yk532d11.5 comes from this gene~cDNA EST yk536g11.5 comes from this gene~cDNA EST yk642c12.5 comes from this gene	222	51
590	ABB12373	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 128.	587	88
590	gi12698103	Macaca fascicularis	hypothetical protein	505	96
590	AAG02711	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6792.	411	97
591	gi14336677	Homo sapiens	16p13.3 sequence section 1 of 8.	673	100
591	gi14327922	Homo sapiens	hypothetical protein FLJ22940, clone MGC:14880 IMAGE:3946937, mRNA, complete cds.	673	100
591	gi12655063	Homo sapiens	polymerase (RNA) III (DNA directed) polypeptide K (12.3 kDa), clone MGC:668 IMAGE:3051476, mRNA, complete cds.	673	100
592	gi9651111	Macaca fascicularis	hypothetical protein	495	74
592	AAO06794	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 20686.	110	37
592	gi3882271	Homo sapiens	mRNA for KIAA0775 protein, complete cds.	101	29
593	gi12848554	Mus musculus	putative	1362	96
593	gi8655657	Homo sapiens	mRNA; cDNA DKFZp762O076 (from clone DKFZp762O076).	1041	100
593	gi12804029	Homo sapiens	clone IMAGE:3940519, mRNA, partial cds.	754	51
594	gi2190184	Homo sapiens	mRNA for zinc finger protein, complete cds.	616	100
594	gi12803507	Homo sapiens	zinc finger protein, clone MGC:717 IMAGE:3143091, mRNA, complete cds.	616	100
594	AAB58863	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 571.	599	97
595	gi15080543	Homo sapiens	Similar to RIKEN cDNA 5031425D22 gene, clone MGC:21579 IMAGE:4473003, mRNA, complete cds.	1254	100
595	AAY35940	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 189.	1051	99
595	gi12860261	Mus musculus	putative	1007	78
596	AAB43377	Homo sapiens	CURA- Human ORFX ORF3141 polypeptide sequence SEQ ID NO:6282.	807	99
596	gi16877603	Homo sapiens	Similar to SNARE Vti1a-beta protein, clone MGC:9292 IMAGE:3885564, mRNA, complete cds.	711	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
596	gi3421062	Mus musculus	29-kDa Golgi SNARE	700	98
597	gi13384259	Homo sapiens	apolipoprotein L6 mRNA, complete cds.	1550	99
597	AAM93925	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 4091.	1341	100
597	gi6562077	Homo sapiens	Human DNA sequence from clone SC22CB-33F2 on chromosome 22 Contains part of the gene for a novel protein similar to C-terminal parts of APOL (apolipoprotein L) and TNF-inducible protein CG12-1. Contains GSSs, complete sequence.	1251	100
598	AAG01189	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5270.	301	98
598	AAM40924	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5855.	106	41
598	ABB11379	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1749.	106	41
599	gi12848031	Mus musculus	putative	504	76
599	gi12718388	Neurospora crassa	conserved hypothetical protein	186	37
599	gi9758240	Arabidopsis thaliana		141	27
600	AAG04048	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8129.	553	100
600	AAM25836	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1351.	501	73
600	ABB15766	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 4423.	365	80
601	AAM25836	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1351.	645	77
601	AAG04048	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8129.	553	100
601	AAG02274	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6355.	276	96
602	gi17133695	Nostoc sp. PCC 7120	WD-40 repeat-protein	65	45
603	gi7243278	Homo sapiens	mRNA for KIAA1440 protein, partial cds.	2003	100
603	gi7291723	Drosophila melanogaster	CG3173 gene product	1815	34
603	gi13279125	Homo sapiens	clone IMAGE:3618123, mRNA, partial cds.	1779	100
604	AAY12244	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 557.	378	87
604	AAY59717	Homo sapiens	GEST Secreted protein 58-49-3-G10-FL1.	378	87
604	gi2291129	Caenorhabditis elegans	Hypothetical protein C02A12.5	78	30
605	gi15074866	Tuber magnatum	protein kinase C homologue	82	32
605	gi7110512	Gallus gallus	TGF-beta signal transducer Smad8	79	37
605	AAM93694	Homo sapiens	HELI- Human polypeptide, SEQ ID	75	63

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 3606.		
606	AAU16929	Homo sapiens	HUMA- Human novel secreted protein, SEQ ID 170.	1118	99
606	AAU17002	Homo sapiens	HUMA- Human novel secreted protein, SEQ ID 243.	1117	100
606	gi13623247	Homo sapiens	Similar to RIKEN cDNA 1110001K21 gene, clone MGC:11275 IMAGE:3944355, mRNA, complete cds.	1082	100
607	gi12698049	Homo sapiens	mRNA for KIAA1752 protein, partial cds.	2706	99
607	gi6103000	Mus musculus	fatso protein	2384	86
607	gi12855822	Mus musculus	putative	463	80
608	AAB93514	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12846.	312	100
608	AAG01489	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5570.	312	100
608	AAW61552	Homo sapiens	ABBO Human endosulfine B protein.	312	100
609	gi15341686	Homo sapiens	clone MGC:20522 IMAGE:4578480, mRNA, complete cds.	1695	100
609	gi14349357	Homo sapiens	hypothetical protein FLJ22501, clone MGC:14897 IMAGE:3939754, mRNA, complete cds.	1695	100
609	gi10438914	Homo sapiens	cDNA: FLJ22501 fis, clone HRC11368.	1695	100
610	AAM93816	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3867.	1051	95
610	gi9280104	Macaca fascicularis	unnamed protein product	1035	48
610	AAE07112	Homo sapiens	HUMA- Human gene 6 encoded secreted protein fragment, SEQ ID NO:129.	1033	49
611	AAG93313	Homo sapiens	NISC- Human protein HP10569.	365	100
611	gi17389971	Homo sapiens	clone IMAGE:4251653, mRNA, partial cds.	365	100
611	AAG02098	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6179.	300	100
612	gi12654899	Homo sapiens	Similar to x 006 protein, clone MGC:5294 IMAGE:3452502, mRNA, complete cds.	1110	100
612	AAB41932	Homo sapiens	CURA- Human ORFX ORF1696 polypeptide sequence SEQ ID NO:3392.	1091	100
612	gi9437345	Homo sapiens	x 006 protein mRNA, complete cds.	1022	97
613	gi11611571	Macaca fascicularis	hypothetical protein	220	89
613	gi9280196	Macaca fascicularis	unnamed protein product	111	34
613	gi12846582	Mus musculus	putative	88	28
614	AAG02925	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7006.	275	96
614	gi402177	Candida albicans	Fatty acid synthase subunit beta	65	41

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
614	gi1592041	Methanococcus jannaschii	conserved hypothetical protein	65	31
615	gi15787978	Homo sapiens	nuclear export factor 3 (NXF3) mRNA, complete cds.	2824	100
615	gi11230440	Homo sapiens	mRNA for nuclear RNA export factor 3 (NXF3 gene).	2824	100
615	gi12053833	Homo sapiens	partial mRNA for nuclear RNA export factor 3 (NXF3 gene).	1794	99
616	gi7770141	Homo sapiens	PRO1728	662	100
616	gi169156	Pisum sativum	ribulose 1,5-bisphosphate carboxylase small subunit propeptide	73	25
616	gi17862888	Drosophila melanogaster	SD01663p	72	31
617	AAY27630	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 64.	220	100
618	gi15487240	Homo sapiens	mRNA for putative autophagy-related cysteine endopeptidase 2 (AUTL2 gene).	2138	99
618	gi4176500	Homo sapiens	Human DNA sequence from clone 889N15 on chromosome Xq22.1-22.3. Contains part of the gene for a novel protein similar to X. laevis Cortical Thymocyte Marker CTX, the possibly alternatively spliced gene for 26S Proteasome subunit p28 (Ankyrin repeat protein), a novel gene and exons 36 through 45 of the COL4A6 for Collagen Alpha 6(IV). Contains ESTs, STSs, GSSs and a putative CpG island, complete sequence.	2123	100
618	gi15487242	Homo sapiens	mRNA for putative autophagy-related cysteine endopeptidase 2, short splice variant (AUTL2 gene).	1446	73
619	gi2558947	Bacillus subtilis	ParC	89	23
619	gi2634193	Bacillus subtilis	DNA gyrase-like protein (subunit A)	88	23
619	gi1405462	Bacillus subtilis	GrlA	88	23
620	gi12583981	Homo sapiens	transmembrane 6 superfamily member 2 (TM6SF2) mRNA, partial cds.	1386	90
620	gi12583979	Homo sapiens	transmembrane 6 superfamily member 1 (TM6SF1) mRNA, complete cds.	830	54
620	AAG89336	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 456.	828	54
621	gi17384428	Homo sapiens	Human DNA sequence from clone RP11-100C15 on chromosome 9q34.2-34.3 Contains the 3' end of a novel gene for a protein similar to KIAA1543 protein, the gene for a novel potassium channel subunit protein (KIAA1422), part of a novel gene, the 5' end of a gene for a novel lipocalin/cytosolic	4928	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			fatty-acid binding protein and CpG islands, complete sequence.		
621	gi15215360	Homo sapiens	clone IMAGE:3939659, mRNA, partial cds.	3270	99
621	gi14714974	Homo sapiens	clone IMAGE:3865907, mRNA, partial cds.	1090	100
622	AAB66590	Homo sapiens	UYBR- Human KARP-1 protein.	932	91
622	gi307094	Homo sapiens	Human Ku (p70/p80) subunit mRNA, complete cds.	923	92
622	gi307093	Homo sapiens	Human Ku autoimmune antigen gene, complete cds.	923	92
623	AAY73468	Homo sapiens	GEMY Human secreted protein clone yd88_1 protein sequence SEQ ID NO:158.	601	91
623	gi7292183	Drosophila melanogaster	CG12361 gene product	75	32
623	gi5911822	Homo sapiens	Human DNA sequence from clone RP3-526I14 on chromosome 22 Contains the BZRP gene for peripheral benzodiazapine receptor (PBR, PKBS, mitochondrial benzodiazepine, MBR), the KIAA0153 gene, and the gene for a novel CUB and EGF-like domains containing protein. Contains ESTs, STSs, GSSs, genomic marker D22S1179, a ca repeat polymorphism and a putative CpG island, complete sequence.	74	33
624	gi15788454	Mus musculus	growth hormone-inducible soluble protein	409	92
624	gi7298358	Drosophila melanogaster	CG6115 gene product	215	50
624	gi7529571	Homo sapiens	Human DNA sequence from clone RP1-122O8 on chromosome 6q14.2-16.1. Contains the 3' part of a novel gene partially coded for by KIAA0301, a novel gene and the 3' part of the gene KIAA0957. Contains ESTs, STSs, GSSs and a putative CpG island, complete sequence.	93	34
625	gi9967224	Macaca fascicularis	hypothetical protein	337	98
625	gi577220	Saccharomyces cerevisiae	Stt4p: Phosphatidylinositol-4-kinase	68	42
625	gi454207	Saccharomyces cerevisiae	homologous protein to PI3-kinase (STT4)	68	42
626	gi7291693	Drosophila melanogaster	CG16787 gene product	233	36
626	gi4966353	Arabidopsis thaliana	ESTs gb|T76348, gb|N65615 and gb|Z18119 come from this gene.	110	26
626	gi17104753	Arabidopsis thaliana	unknown protein	99	26
627	gi12856787	Mus musculus	putative	785	98

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
627	AAG02618	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6699.	319	100
627	gi4218005	Arabidopsis thaliana	putative vicilin storage protein (globulin-like)	101	23
628	gi12834588	Mus musculus	putative	420	65
628	gi7299316	Drosophila melanogaster	CG12816 gene product	99	40
628	AAM83343	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:10936.	82	34
629	AAB50865	Homo sapiens	UNIW Modified human annexin, SEQ ID NO: 6.	565	99
629	AAB50864	Homo sapiens	UNIW Modified human annexin, SEQ ID NO: 4.	565	99
629	AAB50863	Homo sapiens	UNIW Modified human annexin, SEQ ID NO: 2.	565	99
630	AAB50865	Homo sapiens	UNIW Modified human annexin, SEQ ID NO: 6.	163	96
630	AAB50864	Homo sapiens	UNIW Modified human annexin, SEQ ID NO: 4.	163	96
630	AAB50863	Homo sapiens	UNIW Modified human annexin, SEQ ID NO: 2.	163	96
631	AAE04909	Homo sapiens	INCY- Human transporter and ion channel-22 (TRICH-22) protein.	3324	100
631	AAB24281	Homo sapiens	UROG- Prostate tumour associated gene 24P4C12 protein sequence SEQ ID NO:2.	3320	99
631	AAB93981	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14063.	3313	99
632	AAG81401	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:320.	229	100
632	AAG93300	Homo sapiens	NISC- Human protein HP10417.	229	100
632	AAG00912	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4993.	229	100
633	AAG89339	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 459.	861	100
633	gi13397925	Mus musculus	hypothetical protein	815	94
633	gi12850449	Mus musculus	putative	814	94
634	AAB94808	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15947.	708	100
634	gi10436192	Homo sapiens	cDNA FLJ13912 fis, clone Y79AA1000230.	708	100
634	gi15680180	Homo sapiens	clone MGC:22939 IMAGE:4870865, mRNA, complete cds.	404	91
635	gi14091315	Mus musculus	ADMP	371	85
635	gi16877066	Homo sapiens	clone MGC:24447 IMAGE:4077762, mRNA, complete cds.	173	45
635	gi16877059	Homo sapiens	clone MGC:24437 IMAGE:4075637, mRNA, complete cds.	173	45
636	gi10442725	Homo sapiens	pellino related intracellular signalling molecule (PRISM) mRNA, complete cds.	2273	100
636	gi10242359	Homo sapiens	pellino 1 (PELI1) mRNA, complete	2273	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			cds.		
636	gi16741380	Mus musculus	pellino (Drosophila) homolog 1	2268	99
637	gi330178	human herpesvirus 1	ORF1	77	32
637	AAY17406	Homo sapiens	UYHU- Human atrophin-1 related protein.	76	35
637	gi8096340	Homo sapiens	mRNA for RERE, complete cds.	76	35
638	AAB42962	Homo sapiens	CURA- Human ORFX ORF2726 polypeptide sequence SEQ ID NO:5452.	1099	100
638	gi3342738	Homo sapiens	chromosome 19, cosmid R26660, complete sequence.	358	93
638	AAG03426	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7507.	315	100
639	AAY00293	Homo sapiens	HUMA- Human secreted protein encoded by gene 36.	645	86
639	AAM23891	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1416.	394	97
639	AAY12138	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 451.	217	100
640	gi15341790	Homo sapiens	Similar to RIKEN cDNA 2900009I07 gene, clone MGC:17347 IMAGE:2901027, mRNA, complete cds.	1484	100
640	gi12837626	Mus musculus	putative	1414	96
640	AAG74211	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4975.	400	64
641	gi14017855	Homo sapiens	mRNA for KIAA1819 protein, partial cds.	2032	99
641	gi14017849	Homo sapiens	mRNA for KIAA1816 protein, partial cds.	253	25
641	gi6979930	Homo sapiens	Mam1 mRNA, partial cds.	195	24
642	gi10439151	Homo sapiens	cDNA: FLJ22671 fis, clone HSI08712.	1445	100
642	AAE07108	Homo sapiens	HUMA- Human gene 3 encoded secreted protein fragment, SEQ ID NO:125.	881	98
642	AAE07053	Homo sapiens	HUMA- Human gene 3 encoded secreted protein HWHSO13, SEQ ID NO:70.	768	99
643	AAB94047	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14209.	1038	100
643	gi14327927	Homo sapiens	hypothetical protein FLJ12474, clone MGC:15036 IMAGE:3678268, mRNA, complete cds.	1038	100
643	gi10433982	Homo sapiens	cDNA FLJ12474 fis, clone NT2RM1000927.	1038	100
644	AAU00784	Homo sapiens	INCY- Human apoptosis protein, APOP-4.	1941	100
644	gi13544020	Homo sapiens	Similar to RIKEN cDNA 6030457N17 gene, clone MGC:13096 IMAGE:3944994, mRNA, complete cds.	1941	100
644	gi12833947	Mus musculus	putative	1382	69

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
645	gi387048	Cricetus cricetus	DHFR-coamplified protein	1037	85
645	AAU19758	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 408.	538	100
645	AAU21495	Homo sapiens	HUMA- Human novel foetal antigen, SEQ ID NO 1739.	538	100
646	gi16565963	Homo sapiens	SAM-dependent methyltransferase gene, exon 11 and complete cds; and SAM-dependent methyltransferase gene, complete cds, alternatively spliced.	1076	90
646	gi15342055	Homo sapiens	hypothetical protein MGC2454, clone MGC:4132 IMAGE:2961526, mRNA, complete cds.	1076	90
646	gi13278783	Homo sapiens	clone MGC:2454 IMAGE:2961526, mRNA, complete cds.	1076	90
647	AAG03651	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7732.	199	76
647	gi8927662	Unknown	Contains similarity to extensin (atExt1) from Arabidopsis thaliana gb|U43627 and is rich	84	39
647	gi7294152	Drosophila melanogaster	CG13048 gene product	83	41
648	AAY12550	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 215 from WO 9906553.	163	100
648	gi9759124	Arabidopsis thaliana	salt-inducible protein-like	66	37
648	gi15237345	Arabidopsis thaliana] > [Arabidopsis thaliana	salt-inducible protein-like	66	37
649	gi1262852	Mus musculus	M17 protein	413	55
649	gi13874586	Macaca fascicularis	hypothetical protein	150	34
649	gi15150696	Caenorhabditis elegans	Hypothetical protein Y55B1BR.3	80	32
650	gi12862482	Homo sapiens	ALS2CR3 mRNA for amyotrophic lateral sclerosis 2, candidate 3, complete cds.	2969	99
650	gi12862664	Homo sapiens	ALS2CR3 gene for amyotrophic lateral sclerosis 2, candidate 3, exon 16 and complete cds.	2963	99
650	AAY92241	Homo sapiens	LUDW- Human cancer associated antigen precursor (MO-REN-46).	2962	99
651	gi14043592	Homo sapiens	hypothetical protein FLJ13154, clone MGC:13154 IMAGE:4302289, mRNA, complete cds.	1401	100
651	gi13623389	Homo sapiens	hypothetical protein FLJ13154, clone MGC:10683 IMAGE:4025993, mRNA, complete cds.	1401	100
651	gi13325194	Homo sapiens	hypothetical protein FLJ13154, clone MGC:11014 IMAGE:3641317, mRNA, complete cds.	1401	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
652	gi12841092	Mus musculus	putative	1442	90
652	AAB43804	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1249.	531	85
652	gi466475	Geobacillus stearothermophilus	putative phospho-beta-glucosidase	261	33
653	gi16550394	Homo sapiens	cDNA FLJ31056 fis, clone HSYRA2000760.	1412	99
653	gi16648324	Drosophila melanogaster	LD29159p	265	42
653	gi7295644	Drosophila melanogaster	CG14613 gene product	265	42
654	AAY53056	Homo sapiens	GEMY Human secreted protein clone my340_1 protein sequence SEQ ID NO:118.	479	100
655	gi7293719	Drosophila melanogaster	CG14182 gene product	480	51
655	gi16648454	Drosophila melanogaster	SD01285p	79	22
655	gi7291881	Drosophila melanogaster	CG3770 gene product	79	22
656	gi15146320	Arabidopsis thaliana	At2g27260/F12K2.16	79	34
656	gi13272403	Arabidopsis thaliana	unknown protein	79	34
656	gi3608135	Arabidopsis thaliana	putative G-box-binding bZIP transcription factor	74	26
657	gi10439656	Homo sapiens	cDNA: FLJ23082 fis, clone LNG06451.	1960	99
657	AAB95383	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17715.	1222	100
657	gi10435167	Homo sapiens	cDNA FLJ13231 fis, clone OVARC1000145.	1222	100
658	gi17046389	Homo sapiens	C21orf70 isoform B protein (C21orf70) mRNA, complete cds, alternatively spliced.	606	100
658	gi17046387	Homo sapiens	C21orf70 isoform A protein (C21orf70) mRNA, complete cds, alternatively spliced.	606	100
658	gi14424633	Homo sapiens	clone MGC:16722 IMAGE:4128732, mRNA, complete cds.	606	100
659	AAO09511	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 23403.	98	38
659	AAO09309	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 23201.	92	56
659	gi220579	Mus musculus	open reading frame (196 AA)	88	57
660	AAB94146	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14423.	2585	100
660	gi13325430	Homo sapiens	hypothetical protein FLJ12584, clone MGC:11212 IMAGE:3929097, mRNA, complete cds.	2585	100
660	gi10434160	Homo sapiens	cDNA FLJ12584 fis, clone NT2RM4001187.	2585	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
661	AAY59708	Homo sapiens	GEST Secreted protein 76-20-4-C11-FL1.	196	95
661	AAB43261	Homo sapiens	CURA- Human ORFX ORF3025 polypeptide sequence SEQ ID NO:6050.	184	97
661	gi15451283	Macaca fascicularis	hypothetical protein	179	97
662	gi12834045	Mus musculus	putative	309	57
662	AAM79478	Homo sapiens	HYSE- Human protein SEQ ID NO 3124.	306	52
662	AAM78494	Homo sapiens	HYSE- Human protein SEQ ID NO 1156.	306	52
663	AAB87406	Homo sapiens	HUMA- Human gene 32 encoded secreted protein HELHN47, SEQ ID NO:147.	1862	91
663	AAY86456	Homo sapiens	HUMA- Human gene 46-encoded protein fragment, SEQ ID NO:371.	1862	91
663	AAY86260	Homo sapiens	HUMA- Human secreted protein HELHN47, SEQ ID NO:175.	1862	91
664	AAW75222	Homo sapiens	HUMA- Human secreted protein encoded by gene 27 clone H2MBT68.	208	100
664	gi3874864	Caenorhabditis elegans	C38C6.4	70	36
664	gi7497178	Caenorhabditis elegans	hypothetical protein C38C6.4 - Caenorhabditis elegans >	70	36
665	gi9929941	Macaca fascicularis	hypothetical protein	486	89
665	AAM99916	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 32.	70	36
665	gi9929941	Macaca fascicularis	hypothetical protein	486	89
666	gi10438496	Homo sapiens	cDNA: FLJ22202 fis, clone HRC01333.	915	100
666	gi1946267	Oryza sativa	myb	80	31
666	AAB64815	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 43 SEQ ID NO:101.	79	30
667	AAG03788	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7869.	113	34
667	AAM24321	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1846.	107	56
667	AAY65066	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1227.	88	50
668	gi11611585	Macaca fascicularis	hypohtetical protein	1798	90
668	gi12698180	Macaca fascicularis	hypothetical protein	1789	89
668	gi13279047	Homo sapiens	clone MGC:10761 IMAGE:3606108, mRNA, complete cds.	1446	100
669	gi7417266	Homo sapiens	chromosome X map Xp11.23 L-type calcium channel alpha-1 subunit (CACNA1F) gene, complete cds; HSP27 pseudogene, complete	4039	99

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			sequence; and JM1 protein, JM2 protein, and Hb2E genes, complete cds.		
669	gi13559955	Mus musculus	DXImx48e protein	3034	79
669	gi165693	Oryctolagus cuniculus	protein phosphatase regulatory subunit	220	28
670	AAB43283	Homo sapiens	CURA- Human ORFX ORF3047 polypeptide sequence SEQ ID NO:6094.	715	100
670	gi14250579	Homo sapiens	hypothetical protein PP1628, clone MGC:3072 IMAGE:3346334, mRNA, complete cds.	715	100
670	gi10441903	Homo sapiens	clone PP1628 unknown mRNA.	715	100
671	gi15082451	Homo sapiens	clone MGC:20253 IMAGE:4647654, mRNA, complete cds.	1107	98
671	AAB98620	Homo sapiens	SIIAN- IIuman vacuolar H^+-ATPase C subunit 42.	1105	98
671	gi13277864	Mus musculus	Similar to ATPase, H+ transporting, lysosomal (vacuolar proton pump) 42kD	1016	90
672	AAB73533	Homo sapiens	INCY- Human transferase HTFS-40, SEQ ID NO:40.	150	96
672	AAM40557	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5488.	150	96
672	AAM38771	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1916.	150	96
673	AAE12563	Homo sapiens	ISIS- Human CITEDX (HCITEDX) protein.	994	100
673	gi14495276	Homo sapiens	MRG2 gene, complete cds.	994	100
673	gi5002200	Mus musculus	msg1-related protein 2	712	77
674	gi4590448	Leishmania braziliensis	L6 ribosomal protein	80	34
674	AAY30681	Homo sapiens	GENO- Splice variant ZAP-1B protein of the human tumor suppressor gene ZAP-1.	71	60
674	AAY30680	Homo sapiens	GENO- Splice variant ZAP-1A protein of the human tumor suppressor gene ZAP-1.	71	60
675	gi995537	Homo sapiens	H.sapiens gp70 region of endogenous retrovirus erv-4.	707	100
675	gi995542	Homo sapiens	H.sapiens gp70 region of endogenous retrovirus erv-6.	698	99
675	gi995529	Homo sapiens	H.sapiens gp70 region of endogenous retrovirus erv-16.	690	97
676	gi13816301	Sulfolobus solfataricus	Second ORF in transposon ISC1234	86	45
676	gi13815862	Sulfolobus solfataricus	Transposase ISC1234	86	45
676	gi1707705	Sulfolobus solfataricus	orf c06026	86	45
677	gi6470334	Homo sapiens	protein translocase, JM26 protein, UDP-galactose translocator, pim-2 protooncogene homolog pim-2h, and shal-type potassium channel genes,	914	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			complete cds; JM12 protein and transcription factor IGHM enhancer 3 genes, partial cds; and unknown gene, complete sequence.		
677	gi3258629	Homo sapiens	inner mitochondrial membrane translocase Tim17b mRNA, nuclear gene encoding mitochondrial protein, complete cds.	914	100
677	gi3114824	Homo sapiens	mRNA for (JM3) preprotein translocase, complete CDS (clone IMAGE 345224 and LLOXNC01U138D3 (Baylor College)).	914	100
678	gi6470334	Homo sapiens	protein translocase, JM26 protein, UDP-galactose translocator, pim-2 protooncogene homolog pim-2h, and shal-type potassium channel genes, complete cds; JM12 protein and transcription factor IGHM enhancer 3 genes, partial cds; and unknown gene, complete sequence.	852	77
678	gi3258629	Homo sapiens	inner mitochondrial membrane translocase Tim17b mRNA, nuclear gene encoding mitochondrial protein, complete cds.	852	77
678	gi3114824	Homo sapiens	mRNA for (JM3) preprotein translocase, complete CDS (clone IMAGE 345224 and LLOXNC01U138D3 (Baylor College)).	852	77
679	AAB95758	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18678.	685	100
679	gi14042475	Homo sapiens	cDNA FLJ14739 fis, clone NT2RP3002402.	685	100
679	AAG02020	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6101.	480	98
680	AAY48565	Homo sapiens	META- Human breast tumour-associated protein 26.	336	96
680	gi9967248	Macaca fascicularis	hypothetical protein	318	88
680	gi3834384	Homo sapiens	nuclear localization signal containing protein deleted in Velo-Cardio-Facial syndrome (Nlvcf) mRNA, complete cds.	66	32
681	gi10437387	Homo sapiens	cDNA: FLJ21308 fis, clone COL02131.	2600	99
681	AAG73603	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4367.	2016	100
681	gi6102903	Homo sapiens	mRNA; cDNA DKFZp566D244 (from clone DKFZp566D244); partial cds.	1492	68
682	AAO09836	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 23728.	265	100
682	AAU39010	Homo sapiens	GEMY Human secreted protein	265	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			bf377_1.		
682	gi1695241	Caenorhabditis elegans	Hypothetical protein F20D6.8	67	43
683	AAG03386	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7467.	343	98
683	gi16504195	Salmonella enterica subsp. enterica serovar Typhi	hypothetical protein	78	28
683	gi12328592	Heterodoxus macropus	cytochrome b	66	37
684	gi14250495	Homo sapiens	Similar to RIKEN cDNA 0610006H10 gene, clone MGC:9740 IMAGE:3853707, mRNA, complete cds.	1677	100
684	gi15489134	Homo sapiens	RIKEN cDNA 0610006H10 gene, clone MGC:17267 IMAGE:4155233, mRNA, complete cds.	1159	69
684	gi14789807	Mus musculus	RIKEN cDNA 0610006H10 gene	1159	69
685	AAG73989	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4753.	717	100
685	AAB58998	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 706.	717	100
685	AAM89100	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:16693.	247	61
686	AAY04295	Homo sapiens	HUMA- Human secreted protein encoded by gene 3.	478	97
686	gi211447	Gallus gallus	receptor tyrosine kinase	75	35
686	gi1749624	Schizosaccharomyces pombe	similar to Saccharomyces cerevisiae hypothetical 48.0KD protein in CDC28-ARL1 intergenic region precursor, SWISS-PROT Accession Number P38288	69	43
687	AAY02726	Homo sapiens	HUMA- Human secreted protein encoded by gene 77 clone HE2EC79.	158	100
688	gi9967194	Macaca fascicularis	hypothetical protein	269	94
688	gi9948233	Pseudomonas aeruginosa	probable MFS transporter	69	43
688	gi15026548	Clostridium acetobutylicum	Predicted membrane protein	68	32
689	AAY02923	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 99.	235	100
690	AAG73811	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4575.	1099	96
690	gi17028339	Homo sapiens	clone MGC:10198 IMAGE:3909581, mRNA, complete cds.	966	99
690	gi16740631	Mus musculus	Unknown (protein for MGC:27606)	900	90
691	AAG02438	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6519.	360	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
692	gi16553914	Homo sapiens	cDNA FLJ25202 fis, clone REC05350.	2486	87
692	gi13445910	Homo sapiens	radial spoke protein 3 (RSP3) mRNA, complete cds.	1771	86
692	gi16553419	Homo sapiens	cDNA FLJ33093 fis, clone TRACH2000675, weakly similar to RADIAL SPOKE PROTEIN 3.	1566	88
693	gi16553914	Homo sapiens	cDNA FLJ25202 fis, clone REC05350.	2921	99
693	gi13445910	Homo sapiens	radial spoke protein 3 (RSP3) mRNA, complete cds.	2144	100
693	gi13874516	Macaca fascicularis	hypothetical protein	1799	94
694	AAY13135	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 149.	355	100
694	gi16420959	Salmonella typhimurium LT2	regulator for XapA (LysR family)	74	35
694	gi16503639	Salmonella enterica subsp. enterica serovar Typhi	xanthosine operon transcriptional regulator	74	35
695	AAG00152	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4233.	198	100
695	gi14022310	Mesorhizobium loti	hypothetical protein	66	46
696	gi4959568	Homo sapiens	nuclear pore complex interacting protein NPIP (NPIP) mRNA, complete cds.	1742	99
696	gi2342743	Homo sapiens	Human Chromosome 16 BAC clone CIT987SK-A-589H1, complete sequence.	1724	98
696	AAY10915	Homo sapiens	HUMA- Amino acid sequence of a human secreted peptide.	865	98
697	gi4959568	Homo sapiens	nuclear pore complex interacting protein NPIP (NPIP) mRNA, complete cds.	1583	87
697	gi2342743	Homo sapiens	Human Chromosome 16 BAC clone CIT987SK-A-589H1, complete sequence.	1565	87
697	gi3337385	Homo sapiens	Chromosome 16 BAC clone CIT987SK-A-761H5, complete sequence.	886	63
698	gi4959568	Homo sapiens	nuclear pore complex interacting protein NPIP (NPIP) mRNA, complete cds.	1586	92
698	gi2342743	Homo sapiens	Human Chromosome 16 BAC clone CIT987SK-A-589H1, complete sequence.	1573	91
698	AAY10915	Homo sapiens	HUMA- Amino acid sequence of a human secreted peptide.	865	98
699	gi4959568	Homo sapiens	nuclear pore complex interacting protein NPIP (NPIP) mRNA, complete cds.	1503	88
699	gi2342743	Homo sapiens	Human Chromosome 16 BAC clone	1485	87

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			CIT987SK-A-589H1, complete sequence.		
699	gi3337385	Homo sapiens	Chromosome 16 BAC clone CIT987SK-A-761H5, complete sequence.	871	68
700	gi17389867	Homo sapiens	Similar to protein phosphatase 1, regulatory (inhibitor) subunit 1A, clone MGC:24041 IMAGE:4288919, mRNA, complete cds.	572	100
700	gi10198117	Mus musculus	protein phosphatase inhibitor-1	226	49
700	gi7271433	Rattus norvegicus	protein phosphatase inhibitor-1	223	48
701	gi1710282	Homo sapiens	Human clone 23803 mRNA, partial cds.	1899	100
701	gi15215400	Homo sapiens	hypothetical protein MGC4675, clone MGC:2450 IMAGE:2961135, mRNA, complete cds.	458	37
701	gi13278936	Homo sapiens	Similar to RIKEN cDNA 5430432M24 gene, clone MGC:4675 IMAGE:3532660, mRNA, complete cds.	458	37
702	AAB93771	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13481.	1107	100
702	gi10432902	Homo sapiens	cDNA FLJ11608 fis, clone HEMBA1003976.	1107	100
702	gi6599138	Homo sapiens	mRNA; cDNA DKFZp434I036 (from clone DKFZp434I036); partial cds.	86	23
703	AAW89046	Homo sapiens	HUMA- Polypeptide fragment encoded by gene 182.	196	100
703	gi2313995	Helicobacter pylori 26695	lipid A disaccharide synthetase (lpxB)	74	30
703	gi4155351	Helicobacter pylori J99	LIPID-A-DISACCHARIDE SYNTHASE	68	37
704	gi15930206	Homo sapiens	hypothetical protein FLJ12806, clone MGC:9516 IMAGE:3903579, mRNA, complete cds.	1583	99
704	AAB94314	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14787.	1576	99
704	gi10434510	Homo sapiens	cDNA FLJ12806 fis, clone NT2RP2002235.	1576	99
705	AAY64818	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:979.	429	97
705	gi3913990	Mycobacterium smegmatis	ATP-DEPENDENT PROTEASE LA >	66	37
705	gi122240	Rattus norvegicus	RT1 CLASS II HISTOCOMPATIBILITY ANTIGEN, A BETA CHAIN >	66	28
706	AAB95004	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16665.	664	99
706	gi10433328	Homo sapiens	cDNA FLJ11952 fis, clone HEMBB1000831, weakly similar to Homo sapiens breast cancer nuclear receptor-binding auxiliary protein	664	99

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			(BRX) mRNA.		
706	gi10803146	Streptomyces coelicolor	putative regulatory protein	88	42
707	AAG74480	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5244.	2371	99
707	AAB53417	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:957.	2371	99
707	gi15489153	Homo sapiens	hypothetical protein FLJ11896, clone MGC:16887 IMAGE:3858181, mRNA, complete cds.	1729	100
708	gi12862476	Homo sapiens	SIMPLE mRNA for small integral membrane protein of lysosome/late endosome, complete cds.	903	99
708	gi17391332	Mus musculus	LPS-induced TNF-alpha factor	813	86
708	gi6739573	Mus musculus	TBX1 protein	813	86
709	AAG03860	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7941.	425	72
709	gi337508	Homo sapiens	Human ribosomal protein S25 mRNA, complete cds.	425	72
709	gi13436422	Homo sapiens	ribosomal protein S25, clone MGC:4211 IMAGE:2905996, mRNA, complete cds.	425	72
710	AAB63957	Homo sapiens	LUDW- Human prostate cancer associated antigen protein sequence SEQ ID NO:1319.	696	100
710	gi15082563	Homo sapiens	clone MGC:20481 IMAGE:4644158, mRNA, complete cds.	696	100
710	gi12804525	Homo sapiens	clone IMAGE:2823236, mRNA, partial cds.	696	100
711	gi13929452	Homo sapiens	Human DNA sequence from clone RP3-337O18 on chromosome 20q12-13.1. Contains the PLPT gene encoding Phospholipid Transfer Protein, the PPGB gene coding for Lysosomal Protective Protein precursor (EC 3.4.16.5, Cathepsin A, Carboxypeptidase C) and the gene encoding peroxisomal acyl-CoA thioesterase (PTE1, thioesterase II), four novel genes, the gene for a novel protein similar to Drosophila Neuralized (Neu) and the 5' end of an isoform of the TNNC2 gene for fast troponin C2. Contains three CpG islands, ESTs, STSs and GSSs, complete sequence.	3655	100
711	gi16552100	Homo sapiens	cDNA FLJ32079 fis, clone OCBBF2000013.	3645	99
711	AAM70804	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 31110.	936	100
712	AAY60350	Homo sapiens	META- Human normal bladder tissue EST encoded protein 22.	247	90

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
712	gi13883230	Mycobacterium tuberculosis CDC1551	hydrolase, Ama/HipO/HyuC family	65	44
712	gi2894215	Mycobacterium tuberculosis H37Rv	amiB	65	44
'713	AAY94970	Homo sapiens	GEMY Human secreted protein clone dm365_3 protein sequence SEQ ID NO:146.	523	100
713	gi15161741	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_pAT_14p	70	41
713	gi17743430	Agrobacterium tumefaciens str. C58 (Dupont)	conserved hypothetical protein	70	41
714	AAG93310	Homo sapiens	NISC- Human protein HP10561.	1124	97
714	gi12858071	Mus musculus	putative	819	73
714	gi12751094	Homo sapiens	PNAS-124 mRNA, complete cds.	667	99
715	AAM78541	Homo sapiens	HYSE- Human protein SEQ ID NO 1203.	788	86
715	gi15080755	Homo sapiens	ribonuclease P subunit (RPP21) mRNA, complete cds.	788	86
715	gi10439106	Homo sapiens	cDNA: FLJ22638 fis, clone HSI06727.	788	86
716	gi12849817	Mus musculus	putative	679	83
716	AAY57925	Homo sapiens	INCY- Human transmembrane protein HTMPN-49.	670	100
716	gi4926831	Arabidopsis thaliana	T17H7.16	111	30
717	gi9885192	Homo sapiens	Human DNA sequence from clone RP5-881L22 on chromosome 20 Contains ESTs, GSSs, STSs and CpG islands. Contains a gene for a novel protein similar to a trypsin inhibitor and four other genes for novel proteins, complete sequence.	1939	100
717	gi14017764	Mus musculus	CG10671-like	348	35
717	gi14017773	Mus musculus	Cg10671-like	348	35
718	gi7959173	Homo sapiens	mRNA for KIAA1456 protein, partial cds.	1942	99
718	gi16741666	Homo sapiens	clone MGC:16945 IMAGE:3867327, mRNA, complete cds.	1942	99
718	gi7301415	Drosophila melanogaster	CG8968 gene product	270	59
719	AAG75423	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6187.	994	98
719	AAB53454	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:994.	994	98
719	gi12839939	Mus musculus	putative	801	92
720	gi14582152	Xenopus laevis	maxi-K potassium channel alpha subunit Slo	151	100
720	gi5577974	Trachemys	calcium-activated potassium channel	151	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		scripta	isoform thc7		
720	gi2072759	Gallus gallus	calcium-activated potassium channel	151	100
721	AAG03177	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7258.	244	100
722	AAG78876	Homo sapiens	SHAN- Human zinc finger protein 36.	1749	100
722	gi12804829	Homo sapiens	clone MGC:4707 IMAGE:3534541, mRNA, complete cds.	1749	100
722	gi10438507	Homo sapiens	cDNA: FLJ22210 fis, clone HRC01503.	1744	99
723	AAO03397	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 17289.	364	89
723	gi10697002	Homo sapiens	Human DNA sequence from clone RP11-408E5 on chromosome 13q11-12.2 Contains an FSH primary respone homolog 1 (FSHPRH1) pseudogene, two genes for novel proteins, a gene for an orthologue of mouse tubulin alpha 3 (TUBA3) or 7 (TUBA7) and a gene for a novel protein similar to DMPK-like CDC42-binding protein kinase beta (CDC42BPB). Contains ESTs, STSs and GSSs, complete sequence.	330	84
723	AAB42069	Homo sapiens	CURA- Human ORFX ORF1833 polypeptide sequence SEQ ID NO:3666.	282	75
724	gi1045612	Human endogenous retrovirus	pol polyprotein	242	71
724	AAO03158	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 17050.	138	41
724	AAM41750	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6681.	134	37
725	AAW88411	Homo sapiens	UYMA- Acute myeloid leukaemia nuclear matrix associated protein AML-1B.	103	100
725	gi966999	Homo sapiens	Human AML1 mRNA for AML1c protein (alternatively spliced product), complete cds.	103	100
725	gi3153104	Homo sapiens	959 kb contig between AML1 and CBR1 on chromosome 21q22, segment 3/3.	103	100
726	gi10437131	Homo sapiens	cDNA: FLJ21106 fis, clone CAS05176.	1268	99
726	gi7294550	Drosophila melanogaster	CG10982 gene product	294	40
726	gi3875258	Caenorhabditis elegans	waek similarty with bacillus amyloliquefaciens permease IIBC (Swiss Prot accession number P41029)~cDNA EST yk573h3.3 comes from this gene~cDNA EST yk573h3.5 comes from this gene~cDNA EST EMBL:AU109975 comes from this gene~cDNA EST EMBL:AU110906	201	46

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			comes from this gene~cDNA EST EMBL:AU112278 comes from this gene~cDNA EST EMBL:AU110642 comes from this gene~cDNA EST EMBL:AU114810 comes from this gene~cDNA EST EMBL:AU114566 comes from this gene~cDNA EST EMBL:AU116117 comes from this gene~cDNA EST EMBL:AU113930 comes from this gene		
727	AAB97828	Homo sapiens	PFIZ Human G protein-coupled receptor PFI-014 protein sequence SEQ ID NO:2.	195	54
727	AAE06763	Homo sapiens	INCY- Human G-protein coupled receptor-13 (GCREC-13) protein.	174	45
727	gi13384175	Homo sapiens	FKSG46	166	44
728	AAG02577	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6658.	263	98
728	ABB12137	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2507.	261	100
728	gi14334860	Arabidopsis thaliana	putative ATP-dependent Clp protease regulatory subunit CLPX	78	39
729	AAG03340	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7421.	230	97
729	gi15075752	Sinorhizobium meliloti	PROBABLE ADENYLOSUCCINATE SYNTHETASE IMP--ASPARTATE LIGASE PROTEIN	64	34
730	AAG02081	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6162.	565	99
730	AAB65702	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 231.	80	26
730	gi15289906	Oryza sativa	hypothetical protein	72	29
731	gi16549183	Homo sapiens	cDNA FLJ30046 fis, clone 3NB692001719.	1593	100
731	ABB11357	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1727.	1470	93
731	AAG00669	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4750.	600	100
732	gi11611585	Macaca fascicularis	hypohtetical protein	2151	90
732	gi12698180	Macaca fascicularis	hypothetical protein	2142	90
732	gi13279047	Homo sapiens	clone MGC:10761 IMAGE:3606108, mRNA, complete cds.	1446	100
733	AAG03184	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7265.	254	84
734	AAB36365	Homo sapiens	ASAH Human TRAF6 binding protein (T6BP) SEQ ID NO:1.	2317	99
734	gi13435951	Mus musculus	Similar to TAK1-binding protein 2; KIAA0733 protein	610	32
734	AAG64616	Homo sapiens	MATS/ Human TAB2 amino acid sequence.	600	32
735	gi9988100	Homo sapiens	Human DNA sequence from clone	562	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			RP3-467N11 on chromosome 6q16.1-16.3 Contains part of a gene for a novel protein. Contains GSSs, STSs, ESTs and a CpG island, complete sequence.		
735	gi1322280	Mus musculus	unconventional myosin VI	78	24
735	gi12321496	Arabidopsis thaliana	hypothetical protein	75	25
736	gi10437991	Homo sapiens	cDNA: FLJ21816 fis, clone HEP01116.	2205	100
736	gi3253105	Caenorhabditis elegans	Hypothetical protein B0041.7	88	22
736	gi5901659	Caenorhabditis elegans	XNP-1	88	22
737	AAB36671	Homo sapiens	TAKE Human secretory protein TGC-715 SEQ ID NO:11.	406	100
737	AAU12423	Homo sapiens	GETH Human PRO1273 polypeptide sequence.	406	100
737	AAM94192	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 2850.	406	100
738	gi12856120	Mus musculus	putative	781	91
738	gi7292255	Drosophila melanogaster	CG16984 gene product	229	33
738	gi161290	Loligo pealei	kinesin heavy chain	101	31
739	gi10439252	Homo sapiens	cDNA: FLJ22746 fis, clone HUV01174.	1284	99
739	gi16549966	Homo sapiens	cDNA FLJ30707 fis, clone FCBBF2001211.	562	41
739	gi13376148	Homo sapiens	hypothetical protein FLJ22746	1284	99
740	AAY86331	Homo sapiens	HUMA- Human secreted protein HLDCE79, SEQ ID NO:246.	179	100
741	AAB70489	Homo sapiens	SREN- Human hHAIERbs-iso protein sequence SEQ ID NO:7.	1116	91
741	AAM25809	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1324.	1116	91
741	ABB11989	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2359.	1116	91
742	AAB70489	Homo sapiens	SREN- Human hHAIERbs-iso protein sequence SEQ ID NO:7.	835	73
742	AAM25809	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1324.	835	73
742	ABB11989	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2359.	835	73
743	AAG03428	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7509.	389	98
743	AAY59723	Homo sapiens	GEST Secreted protein 60-14-2-H10-FL1.	389	98
743	gi12852865	Mus musculus	putative	295	41
744	AAB95034	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16786.	807	100
744	gi10433444	Homo sapiens	cDNA FLJ12057 fis, clone HEMBB1002068.	807	100
744	gi14715075	Mus musculus	mitotic arrest deficient 1-like 1	85	27
745	AAY13128	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 142.	632	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
745	gi12844331	Mus musculus	putative	509	91
745	AAM25781	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1296.	411	48
746	AAB43357	Homo sapiens	CURA- Human ORFX ORF3121 polypeptide sequence SEQ ID NO:6242.	652	54
746	gi12851679	Mus musculus	putative	640	52
746	AAM38640	Homo sapiens	HUMA- Human colorectal cancer antigen SEQ ID NO: 155.	615	62
747	gi16552467	Homo sapiens	cDNA FLJ32372 fis, clone SALGL1000005.	1067	100
747	gi15278389	Homo sapiens	Similar to hypothetical protein, MGC:7036, clone MGC:4797 IMAGE:3544761, mRNA, complete cds.	1067	100
747	gi13097090	Mus musculus	Unknown (protein for MGC:7036)	750	73
748	AAB64418	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA50.	248	100
748	AAM43637	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 315.	248	100
748	AAM43562	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 240.	248	100
749	gi17512087	Homo sapiens	clone IMAGE:4544931, mRNA, partial cds.	733	100
749	gi15488867	Mus musculus	RIKEN cDNA 2210010N10 gene	596	77
749	gi13905220	Mus musculus	Similar to RIKEN cDNA 2210010N10 gene	591	77
750	gi16553708	Homo sapiens	cDNA FLJ25045 fis, clone CBL03591.	580	76
750	AAB65273	Homo sapiens	GETH Human PRO1287 (UNQ656) protein sequence SEQ ID NO:381.	152	31
750	AAB87561	Homo sapiens	GETH Human PRO1287.	152	31
751	AAE02443	Homo sapiens	CHIL- Human beta-glucuronidase (GUS).	290	77
751	AAW93828	Homo sapiens	CAMB- Human GUS protein fragment.	290	77
751	AAR50092	Homo sapiens	BEHW Humanised anti-CEA sFv fragment-human beta-glucuronidase fusionprotein.	290	77
752	AAY54593	Homo sapiens	INCY- Amino acid sequence of a human transferase designated HUTRAN-3.	2334	100
752	AAB43316	Homo sapiens	CURA- Human ORFX ORF3080 polypeptide sequence SEQ ID NO:6160.	2334	100
752	gi5257221	Mus musculus	protein arginine methyltransferase	2289	98
753	AAB43316	Homo sapiens	CURA- Human ORFX ORF3080 polypeptide sequence SEQ ID NO:6160.	2400	100
753	gi5257221	Mus musculus	protein arginine methyltransferase	2355	98
753	AAY54593	Homo sapiens	INCY- Amino acid sequence of a human transferase designated HUTRAN-3.	2334	100
754	AAG00395	Homo sapiens	GEST Human secreted protein, SEQ ID	268	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 4476.		
754	gi14574333	Caenorhabditis elegans	Hypothetical protein Y41D4B.21	66	30
755	AAY10869	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	129	68
755	gi1170402	Perameles gunnii	SPERM PROTAMINE P1 >	63	32
756	AAB95812	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18806.	1914	100
756	gi12652907	Homo sapiens	clone MGC:2603 IMAGE:3350471, mRNA, complete cds.	1914	100
756	gi10436683	Homo sapiens	cDNA FLJ14264 fis, clone PLACE1002004.	1914	100
757	gi11493710	Homo sapiens	p10-binding protein BITE (BITE) mRNA, complete cds.	3022	99
757	AAB95280	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17491.	3014	99
757	gi10434862	Homo sapiens	cDNA FLJ13036 fis, clone NT2RP3001253, weakly similar to NUF1 PROTEIN.	3014	99
758	AAB41848	Homo sapiens	CURA- Human ORFX ORF1612 polypeptide sequence SEQ ID NO:3224.	559	93
758	gi12861339	Mus musculus	putative	443	74
758	AAY36414	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 7.	442	93
759	gi4128039	Homo sapiens	mRNA for TL132.	994	99
759	AAM38692	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1837.	887	95
759	AAM38691	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1836.	887	95
760	gi12320889	Arabidopsis thaliana	ATP-dependent DNA helicase RecQ, putative	69	35
760	gi17426897	Arabidopsis thaliana	helicase	69	35
760	AAM92379	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 1728.	68	43
761	AAB31473	Homo sapiens	ZYMO Amino acid sequence of a human helical cytokine designated Zalpha33.	924	100
761	AAG93271	Homo sapiens	NISC- Human protein HP10431.	924	100
761	gi14198326	Homo sapiens	Similar to RIKEN cDNA 1810038N03 gene, clone MGC:9890 IMAGE:3868437, mRNA, complete cds.	924	100
762	gi9790624	Homo sapiens	testis-specific kinase substrate (TSKS) gene, complete cds.	3062	100
762	gi11068125	Mus musculus	testis specific serine kinase substrate	2084	81
762	AAM95529	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4187.	785	85
763	gi6502963	Mus musculus	KX antigen	944	43
763	gi12841470	Mus musculus	putative	944	43
763	gi4883433	Homo sapiens	mRNA for membrane transport protein	930	44

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			(XK gene).		
764	AAB95836	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18865.	6026	99
764	gi10436735	Homo sapiens	cDNA FLJ14303 fis, clone PLACE2000132.	6026	99
764	gi14971110	Homo sapiens	mucin 16 (MUC16) mRNA, partial cds.	6023	99
765	gi6807698	Homo sapiens	mRNA; cDNA DKFZp434A1014 (from clone DKFZp434A1014); partial cds.	308	48
765	AAM77697	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 38003.	278	74
765	AAM64969	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37074.	278	74
766	AAB95310	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17554.	551	100
766	gi14794914	Mus musculus	capicua protein	101	32
766	gi12836037	Mus musculus	putative	101	32
767	gi4309887	Homo sapiens	PAC clone RP5-1163J12 from 7q21.2-q31.1, complete sequence.	1047	99
767	AAM73703	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 34009.	136	100
767	AAM61008	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 33113.	136	100
768	gi2664295	Homo sapiens	H.sapiens MDR3 gene, exon1, exon2.	141	100
768	gi307181	Homo sapiens	Human membrane glycoprotein P (mdr3) mRNA, complete cds.	136	100
768	gi1006663	Homo sapiens	H.sapiens mRNA for MDR3 P-glycoprotein.	136	100
769	gi12854186	Mus musculus	putative	1703	88
769	gi5596697	Homo sapiens	Novel human gene mapping to chomosome 22.	818	49
769	gi4493522	Homo sapiens	Human DNA sequence from clone RP3-323M22 on chromosome 22 Contains the 5' part of the PACSIN2 (protein kinase C and casein kinase substrate in neurons 2) gene and a novel gene coding for a protein similar to KIAA0173 and worm tubulin tyrosine ligase, genomic marker D22S418, CA repeat, ESTS, STSs, GSSs and putative CpG islands, complete sequence.	818	49
770	AAB94472	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15137.	1284	100
770	gi10434955	Homo sapiens	cDNA FLJ13096 fis, clone NT2RP3002166.	1284	100
770	AAM66773	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27079.	258	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
771	AAB93902	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13857.	892	95
771	gi10433555	Homo sapiens	cDNA FLJ12147 fis, clone MAMMA1000410.	892	95
771	AAG03840	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7921.	89	56
772	gi13325313	Homo sapiens	Similar to RIKEN cDNA 1500005N04 gene, clone MGC:10325 IMAGE:3936182, mRNA, complete cds.	678	100
772	AAG74090	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4854.	500	97
772	gi12837136	Mus musculus	putative	487	75
773	AAB68986	Homo sapiens	UYJO Human polyamine-modulated factor-1 PMF-1.	832	98
773	gi5737759	Homo sapiens	polyamine modulated factor-1 (PMF1) mRNA, complete cds.	832	98
773	gi5737757	Homo sapiens	polyamine modulated factor-1 (PMF1) gene, exons 2 through 5 and complete cds.	832	98
774	gi10440444	Homo sapiens	mRNA for FLJ00058 protein, partial cds.	696	100
774	gi882260	Homo sapiens	Human chromatin assembly factor-I p60 subunit mRNA, complete cds.	86	28
774	gi7768767	Homo sapiens	genomic DNA, chromosome 21q, section 69/105.	86	28
775	gi10437174	Homo sapiens	cDNA: FLJ21135 fis, clone CAS07262.	1236	99
775	AAO01368	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15260.	158	46
775	gi10645308	Leishmania major	L8453.1	101	27
776	AAB87431	Homo sapiens	HUMA- Human gene 14 encoded secreted protein fragment, SEQ ID NO:172.	883	100
776	AAB87398	Homo sapiens	HUMA- Human gene 14 encoded secreted protein HTEAM34, SEQ ID NO:139.	640	100
776	AAB87355	Homo sapiens	HUMA- Human gene 14 encoded secreted protein HTEAM34, SEQ ID NO:96.	640	100
777	gi13374939	Homo sapiens	Human DNA sequence from clone RP11-204H22 on chromosome 20. Contains part of a novel gene, ESTs, STSs and GSSs, complete sequence.	371	100
777	gi12843034	Mus musculus	putative	362	85
777	AAG02702	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6783.	278	98
778	AAG02713	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6794.	299	100
778	gi6563166	Quiscalus lugubris	NADH dehydrogenase subunit 2	68	38
778	AAW57056	Homo sapiens	CHIL- Class II trans activator (CIITA)	66	44

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			polypeptide.		
779	AAY13108	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 122.	237	100
780	gi14124974	Homo sapiens	Similar to CG12113 gene product, clone IMAGE:3532726, mRNA, partial cds.	4048	100
780	gi14602672	Homo sapiens	Similar to CG12113 gene product, clone IMAGE:3928539, mRNA, partial cds.	2702	100
780	gi14603034	Homo sapiens	clone MGC:16733 IMAGE:4129693, mRNA, complete cds.	2557	100
781	gi17223622	Homo sapiens	ATP-binding cassette A6 mRNA, complete cds.	721	100
781	AAY57954	Homo sapiens	INCY- Human transmembrane protein HTMPN-78.	541	100
781	AAM25936	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1451.	484	100
782	gi8979818	Homo sapiens	Human DNA sequence from clone RP3-447E21 on chromosome 6p12.1-21.1 Contains the 5' end of gene similar to bovine chloride channel protein (p64), a fragment similar to X.laevis Xrel2 protein, a fragment similar to Myelin-associated oligodendrocytic basic protein (MOBP-81), a novel pseudogene, a CpG island, ESTs, STSs and GSSs, complete sequence.	954	100
782	gi14031047	Homo sapiens	CLIC5B mRNA, complete cds.	954	100
782	gi4588530	Bos taurus	chloride channel protein p64	398	46
783	AAY72161	Homo sapiens	BAUG/ Human RNA metabolism protein (RMEP-1).	829	100
783	gi4680653	Homo sapiens	CGI-07 protein mRNA, complete cds.	829	100
783	gi15426434	Homo sapiens	CGI-07 protein, clone MGC:13335 IMAGE:4291797, mRNA, complete cds.	829	100
784	gi7298468	Drosophila melanogaster	CG15164 gene product	413	35
784	gi14026730	Mesorhizobium loti	homoserine kinase	359	28
784	gi15075719	Sinorhizobium meliloti	PUTATIVE AMINOTRANSFERASE PROTEIN	300	27
785	AAM65753	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26059.	661	100
785	AAM53375	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25480.	661	100
785	gi13879308	Mus musculus	centromere autoantigen B	368	30
786	AAY11439	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No 261.	163	100
787	gi9967303	Macaca fascicularis	hypothetical protein	297	96
787	AAM55988	Homo sapiens	MOLE- Human brain expressed single	184	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			exon probe encoded protein SEQ ID NO: 28093.		
787	gi7379384	Neisseria meningitidis Z2491	putative pilus assembly protein	68	36
788	gi15080333	Homo sapiens	clone MGC:20510 IMAGE:4542472, mRNA, complete cds.	1380	100
788	AAB41490	Homo sapiens	CURA- Human ORFX ORF1254 polypeptide sequence SEQ ID NO:2508.	1267	81
788	gi12698051	Homo sapiens	mRNA for KIAA1753 protein, partial cds.	1227	73
789	AAY00277	Homo sapiens	HUMA- Human secreted protein encoded by gene 20.	165	100
789	AAB08450	Homo sapiens	COMP- A human kallikrein-2 (KLK-2) splice variant polypeptide.	75	30
789	gi14574289	Caenorhabditis elegans	Hypothetical protein Y37E11C.1	72	58
790	gi16550493	Homo sapiens	cDNA FLJ31139 fis, clone IMR322001185.	1281	99
790	gi3876588	Caenorhabditis elegans	predicted using Genefinder~cDNA EST yk185a11.3 comes from this gene~cDNA EST yk185a11.5 comes from this gene~cDNA EST yk223d12.5 comes from this gene~cDNA EST yk266b2.5 comes from this gene~cDNA EST yk460f10.5 comes from this gene~cDNA EST yk643b12.3 comes from this gene~cDNA EST yk504b3.5 comes from this gene~cDNA EST yk627c11.5 comes from this gene~cDNA EST yk643b12.5 comes from this gene~cDNA EST yk681b10.3 comes from this gene	239	33
790	gi3880607	Caenorhabditis elegans	cDNA EST yk443f7.5 comes from this gene	109	37
791	gi9837427	Lytechinus variegatus	embryonic blastocoelar extracellular matrix protein precursor	271	44
791	gi17135842	Nostoc sp. PCC 7120	ORF_ID:alr7304~similar to hlyA	121	31
791	gi4566524	Rattus norvegicus	Na+/Ca2+-exchanging protein precursor	120	32
792	gi14250766	Homo sapiens	hypothetical protein FLJ21959, clone MGC:14921 IMAGE:4100186, mRNA, complete cds.	2119	100
792	gi10438183	Homo sapiens	cDNA: FLJ21959 fis, clone HEP05511.	2119	100
792	AAY36034	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 419.	1659	97
793	AAB82316	Homo sapiens	UYCO Human immunoglobulin receptor IRTA3 protein.	491	100
793	gi16033594	Homo sapiens	SH2 domain-containing phosphatase anchor protein 2c mRNA, complete cds, alternatively spliced.	491	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
793	gi16033591	Homo sapiens	SH2 domain-containing phosphatase anchor protein 2b mRNA, complete cds, alternatively spliced.	491	100
794	AAG66841	Homo sapiens	SHAN- Human dihydroorotase 40.	710	99
794	gi12052764	Homo sapiens	mRNA; cDNA DKFZp564O0523 (from clone DKFZp564O0523); complete cds.	703	98
794	ABB12204	Homo sapiens	HYSE- Human HSPC304 homologue, SEQ ID NO:2574.	698	98
795	AAU12298	Homo sapiens	GETH Human PRO9820 polypeptide sequence.	874	98
795	AAH23959_aa 1	Homo sapiens	KYOW Human Klotho cDNA, SEQ ID NO:5.	460	52
795	AAB73618	Homo sapiens	KYOW Human Klotho protein encoded by SEQ ID NO:5.	460	52
796	AAU12298	Homo sapiens	GETH Human PRO9820 polypeptide sequence.	169	100
796	AAB29903	Homo sapiens	HUMA- Human secreted protein BLAST search protein SEQ ID NO: 161.	83	40
796	gi1777770	Cavia porcellus	cytosolic beta-glucosidase	83	40
797	AAY13002	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 16.	222	100
798	AAB65161	Homo sapiens	GETH Human PRO203 (UNQ177) protein sequence SEQ ID NO:30.	1901	100
798	AAY66638	Homo sapiens	GETH Membrane-bound protein PRO203.	1901	100
798	AAB19407	Homo sapiens	CHIR Amino acid sequence of a human secreted protein.	1896	99
799	gi16306705	Homo sapiens	clone MGC:3298 IMAGE:3508400, mRNA, complete cds.	962	100
799	AAY58614	Homo sapiens	INCY- Protein regulating gene expression PRGE-7.	571	69
799	AAM42020	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6951.	87	28
800	AAY48414	Homo sapiens	META- Human prostate cancer-associated protein 111.	191	100
800	gi7293155	Drosophila melanogaster	CG8916 gene product	68	27
801	AAG02085	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6166.	271	100
801	gi16421767	Salmonella typhimurium LT2	DNA biosynthesis; DNA primase	67	34
801	gi16504287	Salmonella enterica subsp. enterica serovar Typhi	DNA primase	67	34
802	AAB93911	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13877.	335	97
802	AAM91037	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ	335	97

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			ID NO:18630.		
802	AAG01519	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5600.	335	97
803	gi10438284	Homo sapiens	cDNA: FLJ22032 fis, clone HEP08743.	1485	99
803	gi14017927	Homo sapiens	mRNA for KIAA1855 protein, partial cds.	1214	93
803	gi4589614	Homo sapiens	mRNA for KIAA0985 protein, complete cds.	140	31
804	AAG00145	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4226.	225	95
804	AAY07867	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 16.	225	95
804	AAW71684	Homo sapiens	INCY- Amino acid sequence of the human tumourigenesis associated protein.	225	95
805	AAB41200	Homo sapiens	CURA- Human ORFX ORF964 polypeptide sequence SEQ ID NO:1928.	694	99
805	gi12855307	Mus musculus	putative	377	91
805	AAG02108	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6189.	333	57
806	AAG02252	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6333.	330	98
806	gi6730714	Arabidopsis thaliana	Unknown protein	68	38
806	gi5729893	Homo sapiens] > [Homo sapiens	A kinase (PRKA) anchor protein 6; A-kinase anchor protein 100	63	47
807	AAB93899	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13848.	3873	99
807	gi14042001	Homo sapiens	cDNA FLJ14464 fis, clone MAMMA1000309.	3873	99
807	gi17512096	Homo sapiens	Similar to hypothetical protein FLJ14464, clone IMAGE:4554168, mRNA, partial cds.	2081	100
808	gi12654201	Homo sapiens	clone IMAGE:3449838, mRNA, partial cds.	621	100
808	gi17068388	Homo sapiens	Similar to hypothetical protein FLJ14775, clone MGC:24018 IMAGE:4105917, mRNA, complete cds.	609	99
808	AAG01516	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5597.	446	98
809	AAB58340	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 678.	942	100
809	ABB11637	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2007.	600	100
809	gi16878257	Homo sapiens	clone MGC:29726 IMAGE:4547604, mRNA, complete cds.	477	52
810	ABB11722	Homo sapiens	HYSE- Human V_segment homologue, SEQ ID NO:2092.	382	59
810	gi1199646	Homo sapiens	Human T cell receptor beta chain (TCRB) mRNA, VDJ region, partial	330	57

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			cds.		
810	gi1864067	Callithrix jacchus	T-cell receptor beta chain	328	56
811	gi10437049	Homo sapiens	cDNA: FLJ21047 fis, clone CAS00253.	797	98
811	gi13880570	Mycobacterium tuberculosis CDC1551	conserved hypothetical protein	79	35
811	gi3261634	Mycobacterium tuberculosis H37Rv	hypothetical protein Rv0976c	79	35
812	gi12838791	Mus musculus	putative	566	76
812	AAG01260	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5341.	338	65
812	gi7297946	Drosophila melanogaster	CG5435 gene product	96	25
813	AAB43507	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:952.	1378	98
813	gi4205084	Homo sapiens	Human WW domain binding protein-1 mRNA, complete cds.	1378	98
813	gi14603081	Homo sapiens	Similar to WW domain binding protein 1, clone MGC:15305 IMAGE:4309279, mRNA, complete cds.	1378	98
814	gi15020649	Homo sapiens	mRNA for hypothetical protein and STS SHGC-2390.	1854	100
814	gi10439232	Homo sapiens	cDNA: FLJ22729 fis, clone HSI15685.	793	100
814	gi14290514	Homo sapiens	hypothetical protein FLJ22729, clone MGC:16790 IMAGE:4184795, mRNA, complete cds.	789	99
815	AAY41454	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 30.	232	93
815	gi3758843	Plasmodium falciparum	hypothetical protein, PFC0820w	71	26
815	gi15025672	Clostridium acetobutylicum	Carbamoylphosphate synthase large subunit	67	33
816	AAG03514	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7595.	189	97
816	gi190508	Homo sapiens	Human PRB4 locus salivary proline-rich protein mRNA, complete cds.	80	30
816	gi15196112	human, peripheral blood leukocytes, subject 'J.J.', Genomic Mutant, 753 nt]. [Homo sapiens	PRB4 (PRB4M PO-)=parotid 'o' protein {exon 3}	80	30
817	AAY65007	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1168.	300	100
817	AAG03529	Homo sapiens	GEST Human secreted protein, SEQ ID	300	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 7610.		
817	gi1932727	Homo sapiens	Human armadillo repeat protein mRNA, complete cds.	64	59
818	AAG01406	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5487.	397	100
818	gi12804657	Homo sapiens	clone IMAGE:3354845, mRNA, partial cds.	397	100
818	gi12841742	Mus musculus	putative	320	76
819	gi10440259	Homo sapiens	cDNA: FLJ23537 fis, clone LNG07690.	1045	100
819	gi48491	Vibrio parahaemolyticus	tryptophan synthase; alpha subunit	74	35
819	gi15155988	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_C_1792p	72	23
820	gi10439767	Homo sapiens	cDNA: FLJ23168 fis, clone LNG09905.	1679	99
820	gi3193250	Caenorhabditis elegans	Hypothetical protein ZK1055.1	122	23
820	gi15290033	Oryza sativa	putative myosin heavy chain-like protein	121	23
821	AAB95117	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17106.	1515	100
821	gi10434031	Homo sapiens	cDNA FLJ12505 fis, clone NT2RM2001699.	1515	100
821	gi6056365	Homo sapiens	chromosome 14 clone 99E15 containing gene for KIAA 1036, complete CDS, complete sequence.	857	57
822	ABB44606	Homo sapiens	SWIT- Human wound healing related polypeptide SEQ ID NO 89.	989	100
822	ABB44607	Homo sapiens	SWIT- Human wound healing related polypeptide SEQ ID NO 90.	876	91
822	ABB44596	Homo sapiens	SWIT- Human wound healing related polypeptide SEQ ID NO 55.	747	100
823	AAB94920	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16368.	760	100
823	gi10432815	Homo sapiens	cDNA FLJ11539 fis, clone HEMBA1002748.	760	100
823	gi11071808	Leishmania major	hypothetical protein P214.45	96	31
824	AAE03641	Homo sapiens	INCY- Human extracellular matrix and cell adhesion molecule-5 (XMAD-5).	1599	100
824	gi15559374	Homo sapiens	clone IMAGE:3628973, mRNA, partial cds.	1599	100
824	AAW54090	Homo sapiens	TEXA Homo sapiens BE123 sequence.	1340	99
825	AAB85771	Homo sapiens	INCY- Human drug metabolizing enzyme (ID No. 3861612CD1).	1587	100
825	gi16877032	Homo sapiens	clone MGC:24011 IMAGE:4091916, mRNA, complete cds.	1573	98
825	AAB73512	Homo sapiens	INCY- Human transferase HTFS-19, SEQ ID NO:19.	773	50

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
826	AAY08477	Homo sapiens	ABBO Human BS274 protein epitope 3.	181	100
826	AAY08476	Homo sapiens	ABBO Human BS274 protein epitope 2.	102	100
826	AAY08478	Homo sapiens	ABBO Human BS274 protein epitope 4.	97	100
827	gi2231329	Ovis aries	bactinecin 11	89	37
827	gi3044086	Myxococcus xanthus	unknown	89	35
827	AAY41496	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 70.	88	37
828	gi11093911	Homo sapiens	Bcl-2 related proline-rich protein (BCL2L12) gene, complete cds, alternatively spliced.	1158	100
828	gi14043469	Homo sapiens	Similar to RIKEN cDNA 5430429M05 gene, clone MGC:13155 IMAGE:4302950, mRNA, complete cds.	1150	99
828	AAW38358	Homo sapiens	APOP- Apoptosis associated protein Bbk.	1141	99
829	gi1054887	Homo sapiens	Human HMGI-C chimeric transcript mRNA, partial cds.	239	68
829	AAG02793	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6874.	197	77
829	AAG74844	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5608.	146	59
830	gi15341178	Homo sapiens	lymphocyte alpha-kinase (LAK) mRNA, complete cds.	472	100
830	AAB56768	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1346.	465	98
830	gi12858085	Mus musculus	putative	412	85
831	gi10436233	Homo sapiens	cDNA FLJ13936 fis, clone Y79AA1000802.	2754	100
831	AAB95616	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18326.	2747	100
831	AAO05842	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 19734.	687	97
832	gi15426492	Homo sapiens	hypothetical protein FLJ21657, clone MGC:14939 IMAGE:3621124, mRNA, complete cds.	1029	93
832	gi10437800	Homo sapiens	cDNA: FLJ21657 fis, clone COL08663.	1025	93
832	gi7292406	Drosophila melanogaster	CG10866 gene product	263	35
833	AAY66151	Homo sapiens	META- Human bladder tumour EST encoded protein 9.	412	98
833	gi6690682	Rhodobacter sphaeroides	Orf173	84	36
833	gi14023427	Mesorhizobium loti	maltose-binding protein component of ABC sugar transporter	78	35
834	AAM25486	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1001.	765	100
834	AAV43605_aa	Homo sapiens	CHIR Human secreted protein 5	352	39

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	1		encoding DNA.		
834	AAY03241	Homo sapiens	SAGA Clone HP10484 of a human secretory signal protein (2).	352	39
835	AAB36599	Homo sapiens	INCY- Human FLEXHT-21 protein sequence SEQ ID NO:21.	1332	100
835	gi4929699	Homo sapiens	CGI-115 protein mRNA, complete cds.	1332	100
835	gi12846260	Mus musculus	putative	1018	74
836	AAY10855	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	185	100
837	gi975846	Bos taurus	immunoglobulin lambda light chain variable region	74	33
837	gi3411264	Emericella nidulans	homeodomain DNA-binding transcription factor	70	58
837	gi7299135	Drosophila melanogaster	Mst85C gene product	69	33
838	gi9948733	Pseudomonas aeruginosa	conserved hypothetical protein	75	40
838	AAB34864	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 11 SEQ ID NO:68.	71	38
838	gi6562167	Homo sapiens	mRNA; cDNA DKFZp564M1916 (from clone DKFZp564M1916); partial cds.	71	33
839	gi10437026	Homo sapiens	cDNA: FLJ21031 fis, clone CAE07336.	663	98
839	gi7188828	Gibberella circinata	histone H3	75	39
839	gi5106126	Aeropyrum pernix	172aa long hypothetical protein	75	40
840	gi10439719	Homo sapiens	cDNA: FLJ23132 fis, clone LNG08559.	2269	100
840	gi14017917	Homo sapiens	mRNA for KIAA1850 protein, partial cds.	2256	99
840	gi13365945	Macaca fascicularis	hypothetical protein	2093	93
841	AAY21589	Homo sapiens	GEMY Human secreted protein (clone BV278-2).	470	100
841	AAW52984	Homo sapiens	GEMY Homo sapiens clone BV278_2 protein.	420	100
841	AAG03462	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7543.	383	98
842	gi16588712	Homo sapiens	P33 mRNA, complete cds.	1284	94
842	gi14334374	Homo sapiens	leucine zipper protein AF5alpha mRNA, complete cds.	1284	94
842	gi14250169	Homo sapiens	Similar to leucine zipper protein FKSG14, clone MGC:14847 IMAGE:3511065, mRNA, complete cds.	1284	94
843	AAB95308	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17550.	1366	99
843	gi10434984	Homo sapiens	cDNA FLJ13114 fis, clone NT2RP3002603.	1366	99
843	AAB40721	Homo sapiens	CURA- Human ORFX ORF485	1286	98

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			polypeptide sequence SEQ ID NO:970.		
844	gi12839493	Mus musculus	putative	714	68
844	AAG01527	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5608.	666	98
844	gi2950243	Hordeum vulgare	extensin	77	31
845	gi10798772	Homo sapiens	mRNA for p53AIP1gamma, complete cds.	579	100
845	gi10798770	Homo sapiens	mRNA for p53AIP1beta, complete cds.	257	100
845	gi10798768	Homo sapiens	mRNA for p53AIP1, complete cds.	257	100
846	AAB73675	Homo sapiens	INCY- Human oxidoreductase protein ORP-8.	620	100
846	gi12841928	Mus musculus	putative	536	84
846	gi15421813	Salmonella enteritidis	putative protein	350	54
847	AAB95773	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18713.	1180	83
847	gi10436616	Homo sapiens	cDNA FLJ14213 fis, clone NT2RP3003572.	1180	83
847	gi14286252	Homo sapiens	Similar to hypothetical protein FLJ14213, clone MGC:16218 IMAGE:3659247, mRNA, complete cds.	681	100
848	gi16552616	Homo sapiens	cDNA FLJ32480 fis, clone SKNMC2001057.	2291	99
848	gi13278954	Homo sapiens	clone IMAGE:3543931, mRNA, partial cds.	1246	100
848	AAB94905	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16300.	1155	99
849	AAB48789	Homo sapiens	HOSP- Human prostate cancer-predisposing protein, CA7 CG04.	73	42
849	AAM40386	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3531.	73	42
849	gi10862762	Homo sapiens	Human DNA sequence from clone RP4-595C2 on chromosome 1q24.1-25.3 Contains ESTs, STSs and GSSs. Contains the 3' part of the gene for two isoforms of the KIAA0351 protein and the gene for angiopoietin Y1, complete sequence.	73	42
850	gi12248877	Oryctolagus cuniculus	mitsugumin72/junctophilin type1	2009	92
850	gi9927301	Mus musculus	junctophilin type 1	1971	91
850	gi9886738	Homo sapiens	JP3 mRNA for junctophilin type3, complete cds.	1475	67
851	gi10334802	Homo sapiens	fanconi anemia protein E (FANCE) mRNA, complete cds.	2735	100
851	gi12850619	Mus musculus	putative	339	50
851	gi5929884	Rattus norvegicus	nucleolin-related protein NRP	103	24
852	AAY59931	Homo sapiens	META- Human myometrium tumour EST encoded protein 11.	398	98
852	AAY59934	Homo sapiens	META- Human myometrium tumour	215	70

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			EST encoded protein 14.		
852	AAY59933	Homo sapiens	META- Human myometrium tumour EST encoded protein 13.	193	94
853	AAY66147	Homo sapiens	META- Human bladder tumour EST encoded protein 5.	365	98
853	gi12833738	Mus musculus	putative	71	52
853	gi3293036	Pseudomonas putida	xcpY	64	24
854	gi10437476	Homo sapiens	cDNA: FLJ21386 fis, clone COL03414.	1645	100
854	gi17028379	Homo sapiens	Similar to hypothetical protein FLJ22792, clone MGC:22933 IMAGE:4905554, mRNA, complete cds.	1537	98
854	gi791119	Saccharomyces cerevisiae	unknown	81	26
855	AAG62621	Homo sapiens	BIOR- Human SNARE protein 25.	1101	100
855	gi9719422	Rattus norvegicus	SNARE Vti1a-beta protein	1062	96
855	gi9719420	Rattus norvegicus	SNARE Vti1a protein	1012	93
856	AAB39312	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 3 SEQ ID NO:61.	315	98
856	AAW88596	Homo sapiens	HUMA- Secreted protein encoded by gene 63 clone HFEBA88.	307	96
857	AAU14106	Homo sapiens	TRIM- Peptide sequence from human c-fos proto-oncoprotein.	243	100
857	AAR53646	Homo sapiens	YEDA c-fos gene product.	243	100
857	gi6518629	Homo sapiens	gene for cellular oncogene c-fos, partial cds.	243	100
858	gi10798770	Homo sapiens	mRNA for p53AIP1beta, complete cds.	449	100
858	gi10798768	Homo sapiens	mRNA for p53AIP1, complete cds.	440	100
858	gi10798772	Homo sapiens	mRNA for p53AIP1gamma, complete cds.	257	100
859	gi17511697	Homo sapiens	hypothetical protein FLJ14950, clone MGC:31757 IMAGE:5013235, mRNA, complete cds.	901	100
859	AAB95526	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18113.	897	99
859	gi14042838	Homo sapiens	cDNA FLJ14950 fis, clone PLACE2000371, weakly similar to TENSIN.	897	99
860	AAG02557	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6638.	297	100
860	AAG89349	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 469.	241	100
860	AAB42657	Homo sapiens	CURA- Human ORFX ORF2421 polypeptide sequence SEQ ID NO:4842.	126	100
861	gi12855891	Mus musculus	putative	173	68
861	gi5360235	Oryctolagus cuniculus	lectin-like oxidized LDL receptor	77	40

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
861	AAY24153	Chimeric - Homo sapiens	NISB Bovine LOX-1 extracellular region/human IgG1 Fc region chimeric protein.	74	36
862	AAY42390	Homo sapiens	GEMY Alternative reading frame amino acid sequence of lv310_7.	615	100
863	gi15278033	Homo sapiens	nuclear LIM interactor-interacting factor, clone MGC:15065 IMAGE:3687816, mRNA, complete cds.	1356	99
863	gi10257410	Homo sapiens	natural resistance-associated macrophage protein 1 (SLC11A1) gene, complete cds, alternatively spliced; and nuclear LIM interactor-interacting factor (NLI-IF) gene, complete cds.	1356	99
863	gi10257407	Homo sapiens	nuclear LIM interactor-interacting factor (NLI-IF) mRNA, complete cds.	1356	99
864	AAG78191	Homo sapiens	SHAN- Human mitochondrial ATPase coupling factor 6-14.	512	98
864	AAG01252	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5333.	332	98
864	gi12861731	Mus musculus	putative	307	64
865	gi2323287	multiple sclerosis associated retrovirus	polyprotein	304	53
865	gi38333	Homo sapiens	Human endogenous retrovirus pHE.1 (ERV9).	260	60
865	gi17432485	porcine endogenous retrovirus	pol	254	47
866	gi16041690	Homo sapiens	hypothetical protein SP192, clone MGC:16819 IMAGE:3909296, mRNA, complete cds.	2544	100
866	gi10503966	Homo sapiens	clone SP192 unknown mRNA.	2544	100
866	gi10437401	Homo sapiens	cDNA: FLJ21319 fis, clone COL02312.	2540	99
867	gi13938183	Homo sapiens	hypothetical protein FLJ23584, clone MGC:14863 IMAGE:3344580, mRNA, complete cds.	1237	100
867	gi10440321	Homo sapiens	cDNA: FLJ23584 fis, clone LNG14307.	1237	100
867	gi3191978	Streptomyces coelicolor A3(2)	putative protein pII uridylyltransferase	84	27
868	gi10438988	Homo sapiens	cDNA: FLJ22558 fis, clone HSI01557.	841	100
868	gi12852764	Mus musculus	putative	88	36
868	gi7688215	Homo sapiens	Human DNA sequence from clone RP4-788L20 on chromosome 20 Contains the HNF3B (hepatocyte nuclear factor 3, beta) gene, a novel gene, ESTs, STSs, GSSs and five CpG islands, complete sequence.	85	35

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
869	AAE02001	Homo sapiens	USSH Human viral IAP-associated factor (VIAF).	1044	86
869	AAB43903	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1348.	1044	86
869	gi12654393	Homo sapiens	clone MGC:3062 IMAGE:3344703, mRNA, complete cds.	1044	86
870	gi13384257	Homo sapiens	apolipoprotein L5 mRNA, complete cds.	2167	98
870	gi6572236	Homo sapiens	Human DNA sequence from clone RP1-41P2 on chromosome 22 Contains the 3' part of the RBM9 gene for RNA binding motif protein 9 and the 3' part of the gene for a novel protein similar to part of APOL (apolipoprotein L) and TNF-inducible protein CG12-1. Contains ESTs, STSs and GSSs, complete sequence.	1614	97
870	gi13384259	Homo sapiens	apolipoprotein L6 mRNA, complete cds.	478	39
871	gi10732650	Homo sapiens	PP3111 mRNA, complete cds.	452	63
871	gi5051823	Amycolatopsis orientalis	putative peptide synthetase	72	30
871	gi2894188	Amycolatopsis orientalis	PCZA363.3	72	30
872	gi10438351	Homo sapiens	cDNA: FLJ22087 fis, clone HEP15918.	3942	100
872	gi10438800	Homo sapiens	cDNA: FLJ22417 fis, clone HRC08579.	3937	99
872	gi13278208	Mus musculus	Similar to hypothetical protein FLJ22087	3410	86
873	AAO10235	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 24127.	810	99
873	gi1244871	Homo sapiens	dioxin receptor repressor (AHRR) gene, exon 12 and complete cds.	784	89
873	gi6330736	Homo sapiens	mRNA for KIAA1234 protein, partial cds.	776	88
874	AAB94957	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16495.	732	100
874	gi10433031	Homo sapiens	cDNA FLJ11715 fis, clone HEMBA1005223.	732	100
874	gi7620533	Bradyrhizobium japonicum	unknown	80	26
875	gi12652943	Homo sapiens	clone MGC:2488 IMAGE:3351245, mRNA, complete cds.	1031	100
875	gi12053307	Homo sapiens	mRNA; cDNA DKFZp434I209 (from clone DKFZp434I209); complete cds.	1031	100
875	gi12846815	Mus musculus	putative	805	78
876	AAG03976	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8057.	457	92
876	gi5922723	Rattus norvegicus	KPL2	73	35
876	gi16604679	Arabidopsis thaliana	putative WD-repeat membrane protein	67	31
877	gi7959931	Homo sapiens	PRO2893	351	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
877	gi7544787	Sus scrofa	glycoprotein ZP1	68	33
877	gi347421	Sus scrofa	zona pellucida glycoprotein	68	33
878	AAM41443	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6374.	287	83
878	AAM39657	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2802.	287	83
878	AAM82707	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:10300.	287	83
879	AAB68986	Homo sapiens	UYJO Human polyamine-modulated factor-1 PMF-1.	749	98
879	gi5737759	Homo sapiens	polyamine modulated factor-1 (PMF1) mRNA, complete cds.	749	98
879	gi5737757	Homo sapiens	polyamine modulated factor-1 (PMF1) gene, exons 2 through 5 and complete cds.	749	98
880	AAY14462	Homo sapiens	HUMA- Human secreted protein encoded by gene 52 clone HFIUR35.	366	98
880	gi6729212	Clostridium botulinum	NTNHA	67	34
880	gi7240602	Clostridium botulinum	progenitor toxin L nontoxic-nonhemagglutinin component	65	34
881	AAB94110	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14346.	3481	99
881	gi10434088	Homo sapiens	cDNA FLJ12542 fis, clone NT2RM4000534.	3481	99
881	AAG02676	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6757.	210	97
882	gi9956045	Homo sapiens	clone CDABP0066 mRNA sequence.	894	94
882	gi3413800	Homo sapiens	Homo sapien mRNA for putative secretory protein, hBET3.	894	94
882	gi2791804	Homo sapiens	bet3 (BET3) mRNA, complete cds.	894	94
883	gi579068	Bacteriophage phi-80	cII gene (AA 1 - 132)	651	99
883	gi12516141	Escherichia coli O157:H7 EDL933	unknown protein encoded within prophage CP-933U	102	36
883	gi13362232	Escherichia coli O157:H7	hypothetical protein	102	36
884	gi7303583	Drosophila melanogaster	CG9005 gene product	78	33
884	gi12861859	Mus musculus	putative	76	32
884	gi10241798	Streptomyces coelicolor	hypothetical protein SCE41.24c	75	33
885	gi17059636	Homo sapiens	Novel human gene mapping to chomosome 22.	2527	99
885	gi14594694	Mus musculus	adiponutrin	1419	67
885	AAY53641	Homo sapiens	CHIR A bone marrow secreted protein designated BMS42.	880	45
886	AAY36025	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 410.	198	94
886	AAY11423	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No 245.	137	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
887	gi687590	Homo sapiens	Human (AF1q) mRNA, complete cds.	431	93
887	gi16307092	Homo sapiens	ALL1-fused gene from chromosome 1q, clone MGC:17309 IMAGE:3878959, mRNA, complete cds.	431	93
887	gi14250081	Homo sapiens	ALL1-fused gene from chromosome 1q, clone MGC:14664 IMAGE:4103485, mRNA, complete cds.	431	93
888	AAG74085	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4849.	286	94
888	gi14043788	Homo sapiens	clone MGC:14288 IMAGE:4135996, mRNA, complete cds.	286	94
888	AAY36036	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 421.	281	92
889	gi16550275	Homo sapiens	cDNA FLJ30968 fis, clone HEART2000411.	1018	98
889	AAM75969	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36275.	661	100
889	AAM63155	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35260.	661	100
890	gi13559062	Homo sapiens	Human DNA sequence from clone RP11-552M11 on chromosome 1. Contains the OVGP1 gene for oviductal glycoprotein 1 (mucin 9, oviductin), three novel genes, the ATP5F1 gene for mitochondrial F0 complex H+ transporting ATP synthase b1, the ADORA3 gene for adenosine A3 receptor and an RPS29 (40S ribosomal protein S29) pseudogene. Contains ESTs, STSs, GSSs and two CpG islands, complete sequence.	667	100
890	AAY59703	Homo sapiens	GEST Secreted protein 47-2-3-G9-FL2.	509	97
890	AAY11473	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No 295.	472	94
891	gi10439198	Homo sapiens	cDNA: FLJ22704 fis, clone HSI12602.	1336	100
891	gi16877288	Homo sapiens	Similar to Hermansky-Pudlak syndrome 3, clone MGC:21006 IMAGE:4415076, mRNA, complete cds.	1191	100
891	gi16552016	Homo sapiens	cDNA FLJ32013 fis, clone NTONG1000033.	1191	100
892	AAB27247	Homo sapiens	INCY- Human EXMAD-25 SEQ ID NO: 25.	2242	100
892	gi13938404	Homo sapiens	clone MGC:1526 IMAGE:2989807, mRNA, complete cds.	1544	99
892	gi15011984	Homo sapiens	bystin mRNA, complete cds.	1532	99
893	AAG03168	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7249.	415	97
893	gi5911457	Pseudomonas	pyochelin synthetase PchF	75	51

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		aeruginosa			
893	gi14286324	Homo sapiens	high-mobility group (nonhistone chromosomal) protein isoforms I and Y, clone MGC:4242 IMAGE:2962998, mRNA, complete cds.	72	39
894	AAB56417	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:995.	890	97
894	AAB08450	Homo sapiens	COMP- A human kallikrein-2 (KLK-2) splice variant polypeptide.	257	52
894	AAY95014	Homo sapiens	ALPH- Human secreted protein vp3_1, SEQ ID NO:68.	228	67
895	gi11034809	Homo sapiens	leucine-zipper protein FKSG13	1914	99
895	gi2674195	Mus musculus	polymerase I-transcript release factor; PTRF	1779	92
895	gi517089	Gallus gallus	leucine zipper protein	1311	72
896	gi12697951	Homo sapiens	mRNA for KIAA1703 protein, partial cds.	1130	98
896	AAB94772	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15858.	1002	99
896	gi10435978	Homo sapiens	cDNA FLJ13839 fis, clone THYRO1000777.	1002	99
897	AAY87322	Homo sapiens	INCY- Human signal peptide containing protein HSPP-99 SEQ ID NO:99.	888	100
897	AAB90648	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO: 191.	871	98
897	AAG03630	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7711.	463	97
898	AAY35953	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 202.	330	98
898	AAY36105	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 490.	319	95
898	AAG00625	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4706.	269	98
899	AAY64868	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1029.	486	97
900	AAG00723	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4804.	304	100
900	gi330019	Hepatitis E virus	structural viral protein	69	54
900	gi418310	Hepatitis E virus	STRUCTURAL PROTEIN 1 >	69	54
901	gi15779204	Homo sapiens	hypothetical protein FLJ12448, clone MGC:22955 IMAGE:4860511, mRNA, complete cds.	1318	100
901	AAB94014	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14138.	1302	99
901	gi10433939	Homo sapiens	cDNA FLJ12448 fis, clone NT2RM1000300.	1302	99
902	AAB94507	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15214.	1220	100
902	gi10435098	Homo sapiens	cDNA FLJ13188 fis, clone NT2RP3004246.	1220	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
902	gi10439526	Homo sapiens	cDNA: FLJ22977 fis, clone KAT11312.	1201	99
903	gi14517331	Homo sapiens	testis-development related NYD-SP20D mRNA, complete cds.	672	98
903	gi14517329	Homo sapiens	testis-development related NYD-SP20C mRNA, complete cds.	672	98
903	gi14039851	Homo sapiens	testes development-related NYD-SP20 mRNA, complete cds.	672	98
904	AAW88724	Homo sapiens	HUMA- Secreted protein encoded by gene 191 clone HJABZ65.	373	98
904	gi13592178	Leishmania major	Serine Threonine Protein Kinase like protein 4	68	36
904	AAO01200	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15092.	66	62
905	gi10439085	Homo sapiens	cDNA: FLJ22624 fis, clone HSI05951.	1749	100
905	gi13938004	Homo sapiens	Similar to hypothetical protein FLJ22624, clone IMAGE:4104833, mRNA, partial cds.	1290	99
905	AAM38631	Homo sapiens	HUMA- Human colorectal cancer antigen SEQ ID NO: 146.	714	98
906	AAW67838	Homo sapiens	HUMA- Human secreted protein encoded by gene 32 clone HLTCJ63.	448	95
906	gi15029372	Homo sapiens	sorbin polypeptide mRNA, complete cds.	80	31
906	gi12860722	Mus musculus	putative	80	30
907	gi12854928	Mus musculus	putative	688	82
907	gi16552651	Homo sapiens	cDNA FLJ32509 fis, clone SMINT1000054.	592	100
907	AAB53906	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:1446.	491	98
908	AAG93309	Homo sapiens	NISC- Human protein HP10560.	598	100
908	gi9954173	Homo sapiens	DNA polymerase delta smallest subunit p12 (POLDS) mRNA, complete cds.	598	100
908	gi12845953	Mus musculus	putative	492	83
909	AAY48253	Homo sapiens	META- Human prostate cancer-associated protein 39.	334	100
909	gi6458749	Deinococcus radiodurans	hypothetical protein	70	38
909	gi1420437	Saccharomyces cerevisiae	ORF YOR181w	66	55
910	AAY48598	Homo sapiens	META- Human breast tumour-associated protein 59.	370	98
910	gi13424450	Caulobacter crescentus	hypothetical protein	68	32
910	gi15833006	Escherichia coli O157:H7] > [Escherichia coli O157:H7	hypothetical protein	66	39
911	gi16553936	Homo sapiens	cDNA FLJ25219 fis, clone STM00503.	667	100
911	gi14250164	Homo sapiens	Similar to RIKEN cDNA 2310030G06 gene, clone MGC:14839 IMAGE:4294167, mRNA, complete cds.	667	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
911	AAG00856	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4937.	488	98
912	AAG01735	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5816.	349	98
912	gi3834380	Rattus norvegicus	intrinsic factor-B12 receptor precursor	67	33
912	gi9968545	Narcissus pseudonarcissus	beta-carotene hydroxylase	65	33
913	gi13279077	Homo sapiens	clone MGC:10820 IMAGE:3613742, mRNA, complete cds.	373	100
913	AAM91638	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:19231.	352	91
913	gi36424	Homo sapiens	Human sec oncogene for SEC protein.	308	84
914	AAM79478	Homo sapiens	HYSE- Human protein SEQ ID NO 3124.	386	54
914	AAM78494	Homo sapiens	HYSE- Human protein SEQ ID NO 1156.	386	54
914	AAB28200	Homo sapiens	CORI- Human xs99.	386	54
915	gi17391253	Homo sapiens	clone MGC:9850 IMAGE:3865616, mRNA, complete cds.	645	100
915	gi15929794	Homo sapiens	Similar to RNA polymerase 1-3 (16 kDa subunit), clone MGC:21099 IMAGE:3847651, mRNA, complete cds.	645	100
915	gi12805135	Mus musculus	Unknown (protein for IMAGE:3591169)	492	78
916	gi12698063	Homo sapiens	mRNA for KIAA1759 protein, partial cds.	3964	99
916	gi12052965	Homo sapiens	mRNA; cDNA DKFZp566M1046 (from clone DKFZp566M1046); complete cds.	3929	97
916	gi10439143	Homo sapiens	cDNA: FLJ22665 fis, clone HSI08219.	3691	99
917	gi12052965	Homo sapiens	mRNA; cDNA DKFZp566M1046 (from clone DKFZp566M1046); complete cds.	4028	99
917	gi12698063	Homo sapiens	mRNA for KIAA1759 protein, partial cds.	3939	98
917	gi10439143	Homo sapiens	cDNA: FLJ22665 fis, clone HSI08219.	3666	97
918	gi9956045	Homo sapiens	clone CDABP0066 mRNA sequence.	270	56
918	gi3413800	Homo sapiens	Homo sapien mRNA for putative secretory protein, hBET3.	270	56
918	gi2791804	Homo sapiens	bet3 (BET3) mRNA, complete cds.	270	56
919	gi13925848	Homo sapiens	kelch-like protein KLHL4c mRNA, complete cds, alternatively spliced.	765	81
919	gi13925845	Homo sapiens	kelch-like protein KLHL4 mRNA, complete cds, alternatively spliced.	765	81
919	gi12697919	Homo sapiens	mRNA for KIAA1687 protein, partial cds.	765	81
920	gi13185301	Homo sapiens	unnamed protein product	871	100
920	gi14043484	Homo sapiens	Similar to RIKEN cDNA 2810021O14 gene, clone MGC:13159	711	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			IMAGE:4303698, mRNA, complete cds.		
920	gi12850457	Mus musculus	putative	702	81
921	gi6649859	Pneumocystis carinii	kexin-like serine endoprotease	71	75
921	AAO05346	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 19238.	70	62
921	gi1780925	human herpesvirus 5	HCMVIRL4 = TRL4	70	61
922	AAB93760	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13446.	1529	100
922	gi10432860	Homo sapiens	cDNA FLJ11577 fis, clone HEMBA1003556.	1529	100
922	gi12856546	Mus musculus	putative	1257	83
923	gi13177760	Homo sapiens	hypothetical protein FLJ21324, clone MGC:4744 IMAGE:3536686, mRNA, complete cds.	1220	99
923	gi10437407	Homo sapiens	cDNA: FLJ21324 fis, clone COL02394.	1217	99
923	AAB43543	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:988.	1216	99
924	gi10438571	Homo sapiens	cDNA: FLJ22257 fis, clone HRC02873.	902	100
924	gi14518442	Caenorhabditis elegans	Hypothetical protein C01G8.9	85	29
924	AAB84577	Homo sapiens	UYEM- Amino acid sequence of a mature human EP2 peptide.	77	37
925	AAB95224	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17350.	837	99
925	gi10434642	Homo sapiens	cDNA FLJ12891 fis, clone NT2RP2004142.	837	99
925	gi11595611	Neurospora crassa	related to U1 SMALL NUCLEAR RIBONUCLEOPROTEIN C	83	50
926	AAE06150	Homo sapiens	HUMA- Human gene 14 encoded secreted protein fragment, SEQ ID NO:212.	837	100
926	AAY87173	Homo sapiens	HUMA- Human secreted protein sequence SEQ ID NO:212.	837	100
926	AAE06151	Homo sapiens	HUMA- Human gene 14 encoded secreted protein fragment, SEQ ID NO:213.	212	100
927	gi7959917	Homo sapiens	PRO2605	816	100
927	gi14603187	Homo sapiens	hypothetical protein PRO2605, clone MGC:19796 IMAGE:3845525, mRNA, complete cds.	642	100
927	AAY66180	Homo sapiens	META- Human bladder tumour EST encoded protein 38.	370	84
928	gi189989	Homo sapiens	Human protein kinase C-L (PRKCL) mRNA, complete cds.	301	72
928	gi56916	Rattus norvegicus	protein kinase	286	67
928	gi220527	Mus musculus	nPKC-eta	286	67
929	AAG03419	Homo sapiens	GEST Human secreted protein, SEQ ID	273	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 7500.		
929	gi2746865	Caenorhabditis elegans	Hypothetical protein T05A8.4	67	34
930	gi13879555	Mus musculus	binder of Rho GTPase 3	612	79
930	gi5731209	Mus musculus	CRIB-containing BORG3 protein	612	79
930	gi12842166	Mus musculus	putative	612	79
931	gi14017917	Homo sapiens	mRNA for KIAA1850 protein, partial cds.	3878	99
931	gi13365945	Macaca fascicularis	hypothetical protein	2320	94
931	gi10439719	Homo sapiens	cDNA: FLJ23132 fis, clone LNG08559.	2256	99
932	gi10440377	Homo sapiens	mRNA for FLJ00024 protein, partial cds.	937	99
932	gi10440377	Homo sapiens	FLJ00024 protein	937	99
933	gi15207959	Macaca fascicularis	hypothetical protein	632	88
933	gi552009	Streptococcus pyogenes	peptidase	97	25
933	gi13623022	Streptococcus pyogenes M1 GAS	C5A peptidase precursor	95	24
934	gi12860619	Mus musculus	putative	609	96
934	AAM74162	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 34468.	182	97
934	AAG03513	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7594.	136	96
935	AAW88598	Homo sapiens	HUMA- Secreted protein encoded by gene 65 clone HFVHY45.	400	100
935	gi12862020	Mus musculus	putative	269	92
935	AAW88821	Homo sapiens	HUMA- Polypeptide fragment encoded by gene 65.	148	100
936	AAY60495	Homo sapiens	META- Human normal bladder tissue EST encoded protein 167.	326	98
937	AAG81401	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:320.	551	100
937	AAG93300	Homo sapiens	NISC- Human protein HP10417.	551	100
937	AAB43646	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1091.	551	100
938	AAY17388	Homo sapiens	INCY- Human vesicle membrane protein-like protein 1.	950	90
938	gi9802048	Homo sapiens	hypothetical protein SBBI10 mRNA, complete cds.	950	90
938	gi8745394	Homo sapiens	Alu co-repressor 1 (ACR1) mRNA, complete cds.	950	90
939	AAG78658	Homo sapiens	BODE- Human peroxidase antioxidising enzyme 24.	303	60
939	AAG04043	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8124.	303	60
939	AAY17388	Homo sapiens	INCY- Human vesicle membrane protein-like protein 1.	303	60

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
940	AAY17388	Homo sapiens	INCY- Human vesicle membrane protein-like protein 1.	805	79
940	gi9802048	Homo sapiens	hypothetical protein SBBI10 mRNA, complete cds.	805	79
940	gi8745394	Homo sapiens	Alu co-repressor 1 (ACR1) mRNA, complete cds.	805	79
941	gi17132972	Nostoc sp. PCC 7120	ORF_ID:all3838~similar to kinesin light chain	100	25
941	gi1335276	Homo sapiens	Human PRB3 gene (PRB3S) for G1 protein, exon 3.	94	24
941	gi1335274	Homo sapiens	Human prb1 gene for salivary proline-rich protein, exon 3.	93	22
942	AAY22155	Homo sapiens	SAKA/ Human Nck associated protein 1.	3552	59
942	gi4760464	Homo sapiens	mRNA for Nck-associated protein 1 (Nap1), complete cds.	3552	59
942	gi15929137	Homo sapiens	NCK-associated protein 1, clone MGC:8981 IMAGE:3907646, mRNA, complete cds.	3552	59
943	gi54004	Mus musculus	put. RP2 protein (aa 1-357)	1210	63
943	gi7298591	Drosophila melanogaster	CG10194 gene product	472	34
943	gi7298588	Drosophila melanogaster	CG10195 gene product	381	31
944	gi17389434	Homo sapiens	hypothetical protein FLJ22639, clone MGC:22172 IMAGE:4700838, mRNA, complete cds.	876	100
944	gi10439108	Homo sapiens	cDNA: FLJ22639 fis, clone HSI06816.	876	100
944	AAG98701	Homo sapiens	COGE- Human cell death protective cDNA clone CNI-00717 ORF5 protein, SEQ:194.	72	28
945	AAB95692	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18510.	1163	100
945	gi10436474	Homo sapiens	cDNA FLJ14100 fis, clone MAMMA1000855.	1163	100
945	gi7020531	Homo sapiens	cDNA FLJ20433 fis, clone KAT03767.	75	25
946	AAB15389	Homo sapiens	TOYJ Human interleukin 6 receptor protein.	86	26
946	gi4699964	Homo sapiens	PAC clone RP5-953A4 from 7q11.23-q21.1, complete sequence.	85	25
946	gi896310	Mamestra brassicae nucleopolyhedrovirus	unknown protein	84	32
947	AAY12607	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 272 from WO 9906553.	395	98
947	gi17223776	Mus musculus	MLLT6	76	33
947	gi7297961	Drosophila melanogaster	nub gene product	71	34
948	gi17046389	Homo sapiens	C21orf70 isoform B protein (C21orf70) mRNA, complete cds, alternatively spliced.	695	71
948	gi17046387	Homo sapiens	C21orf70 isoform A protein (C21orf70)	670	66

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			mRNA, complete cds, alternatively spliced.		
948	gi14424633	Homo sapiens	clone MGC:16722 IMAGE:4128732, mRNA, complete cds.	670	66
949	gi15779053	Homo sapiens	Similar to RIKEN cDNA 6720467C03 gene, clone MGC:26639 IMAGE:4826612, mRNA, complete cds.	869	100
949	gi12859857	Mus musculus	putative	777	88
949	AAG02322	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6403.	630	99
950	AAG89289	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 409.	374	98
950	AAY45307	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 15.	374	98
950	gi6523815	Homo sapiens	phosphotidylethanolamine N-methyltransferase (PNMT) mRNA, complete cds.	374	98
952	AAB94360	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14887.	3208	99
952	gi10434636	Homo sapiens	cDNA FLJ12888 fis, clone NT2RP2004081.	3208	99
952	gi12855328	Mus musculus	putative	2247	72
953	gi476224	Homo sapiens	Human anion exchanger 3 cardiac isoform (cAE3) mRNA, partial cds.	399	100
953	gi10953762	Mus musculus	anion exchanger 3 cardiac isoform	383	64
953	gi202771	Rattus rattus	ORF-cardiac specific 5' coding region; putative	233	63
954	gi12850828	Mus musculus	putative	173	75
954	gi203519	Rattus norvegicus	cytochrome c oxidase subunit VIc	172	72
954	AAM23875	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1400.	161	70
955	AAY36057	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 442.	313	88
955	AAY35931	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 180.	295	100
955	AAY11851	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No: 451.	171	77
956	gi16549966	Homo sapiens	cDNA FLJ30707 fis, clone FCBBF2001211.	2757	99
956	AAM77437	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 37743.	658	100
956	AAM64659	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36764.	658	100
957	gi16551351	Homo sapiens	cDNA FLJ31509 fis, clone NT2RI1000016.	1226	100
957	gi14133227	Homo sapiens	mRNA for KIAA0970 protein, partial cds.	938	98
957	AAG02178	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6259.	738	98

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
958	gi3800830	Rattus norvegicus	putative four repeat ion channel	711	83
958	gi17901375	Homo sapiens	unnamed protein product	711	83
958	gi7292976	Drosophila melanogaster	CG1517 gene product	382	44
959	AAY60063	Homo sapiens	META- Human endometrium tumour EST encoded protein 123.	235	97
959	AAY60064	Homo sapiens	META- Human endometrium tumour EST encoded protein 124.	231	97
959	gi15081715	Arabidopsis thaliana	At2g41840/T11A7.6	81	36
960	gi13177691	Homo sapiens	Similar to RIKEN cDNA 2410047I02 gene, clone MGC:2560 IMAGE:2989772, mRNA, complete cds.	689	100
960	gi12858411	Mus musculus	putative	585	86
960	AAG01650	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5731.	270	98
961	gi7981304	Homo sapiens	Human DNA sequence from clone RP4-551D2 on chromosome 20q13.2-13.33 Contains the gene for a cadherin-like protein VR20, a novel gene, the PPP1R6 gene for protein phosphatase 1 regulatory subunit 6, the 5' end of the SYCP2 gene for synaptonemal complex protein 2, ESTs, STSs, GSSs and two putative CpG islands, complete sequence.	715	98
961	AAU18881	Homo sapiens	HUMA- Novel prostate gland antigen, Seq ID No 180.	652	100
961	AAM96033	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4691.	652	100
962	gi9622236	Homo sapiens	cadherin-like protein VR20 mRNA, partial cds.	1235	92
962	gi12743872	Homo sapiens	Human DNA sequence from clone RP4-551D2 on chromosome 20q13.2-13.33 Contains the gene for a cadherin-like protein VR20, a novel gene, the PPP1R6 gene for protein phosphatase 1 regulatory subunit 6, the 5' end of the SYCP2 gene for synaptonemal complex protein 2, ESTs, STSs, GSSs and two putative CpG islands, complete sequence.	1235	92
962	AAB47329	Homo sapiens	CURA- FCTR6.	1091	84
963	gi9622236	Homo sapiens	cadherin-like protein VR20 mRNA, partial cds.	1264	100
963	gi12743872	Homo sapiens	Human DNA sequence from clone RP4-551D2 on chromosome 20q13.2-13.33 Contains the gene for a cadherin-like protein VR20, a novel gene, the PPP1R6 gene for protein phosphatase 1 regulatory subunit 6, the 5' end of the	1264	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SYCP2 gene for synaptonemal complex protein 2, ESTs, STSs, GSSs and two putative CpG islands, complete sequence.		
963	AAB47329	Homo sapiens	CURA- FCTR6.	1085	84
964	AAY60064	Homo sapiens	META- Human endometrium tumour EST encoded protein 124.	330	98
965	gi14517637	Homo sapiens	mRNA for RGPR-p117, complete cds.	807	79
965	gi14318616	Homo sapiens	clone MGC:17455 IMAGE:3448742, mRNA, complete cds.	807	79
965	AAG02383	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6464.	505	96
966	AAB94865	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16066.	910	99
966	AAM94039	Homo sapiens	HELI- Human stomach cancer expressed polypeptide SEQ ID NO 149.	910	99
966	gi14718862	Homo sapiens	chronic myelogenous leukemia tumor antigen 66 mRNA, complete cds, alternatively spliced.	910	99
967	AAG02669	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6750.	349	100
968	gi10438452	Homo sapiens	cDNA: FLJ22170 fis, clone HRC00652.	2870	100
968	AAB41640	Homo sapiens	CURA- Human ORFX ORF1404 polypeptide sequence SEQ ID NO:2808.	2037	100
968	gi15928410	Mus musculus	Similar to hypothetical protein FLJ22170	1880	69
970	gi10440460	Homo sapiens	mRNA for FLJ00066 protein, partial cds.	655	99
970	gi4512671	Arabidopsis thaliana	En/Spm-like transposon protein	91	30
970	gi4929130	Arabidopsis thaliana	protodermal factor 1	91	30
971	gi15930209	Homo sapiens	hypothetical protein FLJ22477, clone MGC:9527 IMAGE:3917274, mRNA, complete cds.	882	100
971	gi10438882	Homo sapiens	cDNA: FLJ22477 fis, clone HRC10815.	882	100
971	gi12838990	Mus musculus	putative	156	76
972	AAB94173	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14480.	1542	100
972	gi15215287	Homo sapiens	hypothetical protein FLJ12610, clone MGC:15029 IMAGE:4026495, mRNA, complete cds.	1542	100
972	gi10434201	Homo sapiens	cDNA FLJ12610 fis, clone NT2RM4001565.	1542	100
973	AAB94173	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14480.	1419	93
973	gi15215287	Homo sapiens	hypothetical protein FLJ12610, clone MGC:15029 IMAGE:4026495, mRNA, complete cds.	1419	93

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
973	gi10434201	Homo sapiens	cDNA FLJ12610 fis, clone NT2RM4001565.	1419	93
974	AAY41352	Homo sapiens	HUMA- Human secreted protein encoded by gene 45 clone HTXFH55.	300	100
974	gi15029737	Mus musculus	complement component 2 (within H-2S)	67	58
974	gi192435	Mus musculus	complement component C2	67	58
975	AAB95342	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17623.	721	100
975	gi10435060	Homo sapiens	cDNA FLJ13162 fis, clone NT2RP3003625.	721	100
975	gi7302554	Drosophila melanogaster	CG15094 gene product	79	33
976	AAY65192	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1353.	206	100
977	AAG00539	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4620.	420	93
977	gi7243272	Homo sapiens	mRNA for KIAA1437 protein, partial cds.	199	55
977	gi5824508	Caenorhabditis elegans	contains similarity to Pfam domain: PF00018 (SH3 domain), Score=15.4, E-value=0.00062, N=1~cDNA EST yk300d7.3 comes from this gene~cDNA EST yk300d7.5 comes from this gene~cDNA EST yk310d10.3 comes from this gene~cDNA EST yk310d10.5 comes from this gene~cDNA EST yk553a4.3 comes from this gene~cDNA EST yk553a4.5 comes from this gene~cDNA EST yk622f8.3 comes from this gene~cDNA EST yk622f8.5 comes from this gene~cDNA EST yk674e4.3 comes from this gene~cDNA EST yk674e4.5 comes from this gene	68	33
978	AAM41583	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6514.	620	100
978	AAM39797	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2942.	620	100
978	AAG04036	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8117.	607	99
979	AAY65244	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1405.	207	100
979	AAG00117	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4198.	77	55
979	gi16878268	Homo sapiens	Similar to apolipoprotein L, clone MGC:29731 IMAGE:4661222, mRNA, complete cds.	77	55
980	AAG02124	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6205.	334	100
980	gi4469399	Mus musculus	epithelial sodium channel alpha subunit	69	37
980	gi2148928	Rattus norvegicus	epithelial sodium channel alpha subunit	69	37

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
981	gi7768773	Homo sapiens	genomic DNA, chromosome 21q, section 97/105.	1065	99
981	gi1279678	Saccharomyces cerevisiae	unknown	135	40
981	gi1431023	Saccharomyces cerevisiae	ORF YDL038c	135	40
982	AAG67395	Homo sapiens	SUGE- Amino acid sequence of human protein kinase SGK258.	1687	100
982	AAE00669	Homo sapiens	HUMA- Human protein tyrosine kinase receptor (PTK) from clone HDPSB68.	1679	99
982	gi14017797	Homo sapiens	mRNA for KIAA1790 protein, partial cds.	1679	99
983	gi10440430	Homo sapiens	mRNA for FLJ00050 protein, partial cds.	1433	100
983	AAY84901	Homo sapiens	INCY- A human proliferation and apoptosis related protein.	258	40
983	gi12053225	Homo sapiens	mRNA; cDNA DKFZp434P2235 (from clone DKFZp434P2235); complete cds.	257	40
984	AAG03251	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7332.	153	85
984	gi12859308	Mus musculus	putative	99	65
984	gi7296664	Drosophila melanogaster	CG10981 gene product	68	34
985	AAY12780	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:370.	203	100
985	gi13879614	Mycobacterium tuberculosis CDC1551	PE_PGRS family protein	111	43
985	gi9954108	Trypanosoma cruzi	RNA binding protein RGGm	104	38
986	AAG67032	Homo sapiens	SHAN- Human endothelial monocyte activating polypeptide II-62.	2496	99
986	gi10438461	Homo sapiens	cDNA: FLJ22175 fis, clone HRC00773.	1186	100
986	gi14250826	Homo sapiens	hypothetical protein FLJ22175, clone MGC:14955 IMAGE:4301828, mRNA, complete cds.	1171	99
987	AAB94225	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14591.	1027	100
987	AAB56999	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1577.	1027	100
987	gi10434297	Homo sapiens	cDNA FLJ12666 fis, clone NT2RM4002256.	1027	100
988	AAG03612	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7693.	302	100
988	gi15981929	Yersinia pestis	putative iron ABC transporter, ATP-binding protein	64	32
988	gi16124148	Yersinia pestis] > [Yersinia pestis	putative iron ABC transporter, ATP-binding protein	64	32
989	AAG03478	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7559.	190	83

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
989	gi11360118	Homo sapiens	hypothetical protein DKFZp434M1123.1 - human (fragment) >	63	36
990	AAG73521	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4285.	406	98
990	AAY00280	Homo sapiens	HUMA- Human secreted protein encoded by gene 23.	359	98
991	AAG93309	Homo sapiens	NISC- Human protein HP10560.	339	100
991	gi9954173	Homo sapiens	DNA polymerase delta smallest subunit p12 (POLDS) mRNA, complete cds.	339	100
991	gi12845953	Mus musculus	putative	288	84
992	AAW89035	Homo sapiens	HUMA- Polypeptide fragment encoded by gene 171.	159	100
992	gi5852085	Oryza sativa	zwh0008.1	93	27
992	AAB64815	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 43 SEQ ID NO:101.	87	30
993	gi14717079	Homo sapiens	Human DNA sequence from clone RP3-469A13 on chromosome 20 Contains part of the gene for KIAA0889 and a novel protein similar to KIAA0802, a novel gene, the 5' end of the part of the gene for a novel protein similar to N-myc downstream regulated (NDRG1), ESTs, STSs, GSSs and four CpG islands, complete sequence.	1365	99
993	AAB94598	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15418.	1055	96
993	gi10435333	Homo sapiens	cDNA FLJ13346 fis, clone OVARC1002107.	1055	96
994	AAG02845	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6926.	273	100
995	AAM93342	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2883.	283	60
995	gi9279975	Homo sapiens	mRNA for Reprimo, complete cds.	283	60
995	gi12804111	Homo sapiens	candidate mediator of the p53-dependent G2 arrest, clone MGC:11260 IMAGE:3942270, mRNA, complete cds.	283	60
996	gi2633213	Bacillus subtilis	yhzB	79	35
996	gi9802541	Arabidopsis thaliana	F17L21.23	74	24
996	gi7303166	Drosophila melanogaster	CG12864 gene product	74	33
997	ABB12196	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2566.	424	98
997	AAG03905	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7986.	173	59
997	gi14043862	Homo sapiens	clone MGC:14138 IMAGE:3948518, mRNA, complete cds.	173	59
998	AAM78349	Homo sapiens	HYSE- Human protein SEQ ID NO	72	42

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			1011.		
998	AAM79333	Homo sapiens	HYSE- Human protein SEQ ID NO 2979.	71	41
998	gi15042611	Homo sapiens	Ser/Thr protein kinase PAR-1Balpha mRNA, complete cds.	71	41
999	gi16550716	Homo sapiens	cDNA FLJ31318 fis, clone LIVER1000433, moderately similar to Homo sapiens mRNA for neuropathy target esterase.	2201	100
999	AAM25456	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:971.	1423	100
999	AAY70474	Homo sapiens	INCY- Human cyclic nucleotide-associated protein-2 (CNAP-2).	1422	65
1000	gi7293162	Drosophila melanogaster	CG15603 gene product	71	46
1000	gi4586294	Rhodococcus sp. CIR2	transposase	71	47
1000	gi7300412	Drosophila melanogaster	CG14304 gene product	69	41
1001	AAY07759	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 16.	793	88
1001	gi14603397	Homo sapiens	mitochondrial ribosomal protein S28, clone MGC:19500 IMAGE:4331173, mRNA, complete cds.	787	86
1001	gi4454702	Homo sapiens	HSPC007	787	86
1002	gi16549918	Homo sapiens	cDNA FLJ30671 fis, clone FCBBF1000687, moderately similar to Mus musculus Rap2 interacting protein 8 (RPIP8) mRNA.	1527	95
1002	AAB42726	Homo sapiens	CURA- Human ORFX ORF2490 polypeptide sequence SEQ ID NO:4980.	1314	98
1002	gi2588624	Homo sapiens	BAC clone CTB-60N22 from 7q21, complete sequence.	1314	98
1003	gi10439134	Homo sapiens	cDNA: FLJ22659 fis, clone HSI07953.	756	100
1003	AAM70124	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30430.	157	96
1003	gi14027542	Mesorhizobium loti	hypothetical protein	72	32
1004	AAG62909	Homo sapiens	KLEE/ Amino acid sequence of a human xylosylytransferase (XT).	3614	99
1004	gi11322268	Homo sapiens	partial mRNA for xylosyltransferase I (XT-I gene).	3614	99
1004	gi15209651	Homo sapiens	human XT-I (not completely)	3614	99
1005	AAG02478	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6559.	380	100
1005	AAY86496	Homo sapiens	HUMA- Human gene 61-encoded protein fragment, SEQ ID NO:411.	69	35
1005	AAY86324	Homo sapiens	HUMA- Human secreted protein HSRGW16, SEQ ID NO:239.	69	35
1006	AAB90708	Homo sapiens	GEMY Human CJ397_1 protein sequence SEQ ID 109.	241	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1006	AAW48809	Homo sapiens	GEMY Homo sapiens clone CJ397_1 protein.	241	100
1006	gi671656	Sorghum bicolor	gamma-kafirin preprotein	83	32
1007	AAY59661	Homo sapiens	GEST Secreted protein 108-004-5-0-B7-FL.	408	100
1007	gi431033	Homo sapiens	Human beta-1,4 N-acetylgalactosaminyltransferase mRNA, complete cds.	65	45
1007	gi8250584	Streptomyces coelicolor A3(2)	putative integral membrane protein	65	45
1008	AAG73798	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4562.	653	98
1008	AAG03987	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8068.	653	98
1008	AAB54311	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:763.	653	98
1009	AAB24198	Homo sapiens	HONJ/ Human activation-induced cytidine deaminase SEQ ID NO:8.	1086	100
1009	gi9988410	Homo sapiens	AID mRNA for activation-induced cytidine deaminase, complete CDS.	1086	100
1009	gi9988408	Homo sapiens	AID gene for activation-induced cytidine deaminase, complete cds.	1086	100
1010	gi10439796	Homo sapiens	cDNA: FLJ23189 fis, clone LNG12061.	1172	100
1010	AAM70456	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30762.	467	100
1010	gi2627231	Bos taurus	NDP52	101	28
1011	gi10438050	Homo sapiens	cDNA: FLJ21858 fis, clone HEP02301.	744	97
1011	AAG66887	Homo sapiens	SHAN- Human zinc finger protein 17.	156	30
1011	gi16553140	Homo sapiens	cDNA FLJ32873 fis, clone TESTI2003998, weakly similar to T-CELL RECEPTOR BETA CHAIN ANA 11.	146	38
1012	AAG03653	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7734.	425	100
1012	AAU19393	Homo sapiens	PHAA Human G protein-coupled receptor nGPCR-2326.	87	36
1013	gi6180179	Homo sapiens	transcription factor IGHM enhancer 3, JM11 protein, JM4 protein, JM5 protein, T54 protein, JM10 protein, A4 differentiation-dependent protein, triple LIM domain protein 6, and synaptophysin genes, complete cds; and L-type calcium channel alpha-1 subunit gene, partial cds, complete sequence.	3632	99
1013	gi14250618	Homo sapiens	clone MGC:2962 IMAGE:3139519, mRNA, complete cds.	3077	94
1013	gi7242943	Homo sapiens	mRNA for KIAA1294 protein, partial	297	32

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			cds.		
1014	AAY65004	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1165.	194	79
1014	gi3875122	Caenorhabditis elegans	C50F4.4	68	38
1014	gi7497774	Caenorhabditis elegans	hypothetical protein C50F4.4 - Caenorhabditis elegans >	68	38
1015	AAY60578	Homo sapiens	META- Human normal bladder tissue EST encoded protein 250.	477	100
1016	gi12849116	Mus musculus	putative	1072	76
1016	AAB50970	Homo sapiens	GETH Human PRO4302 protein.	306	35
1016	AAU12446	Homo sapiens	GETH Human PRO4302 polypeptide sequence.	306	35
1017	gi2313745	Helicobacter pylori 26695	H. pylori predicted coding region HP0614	73	35
1017	gi10038760	Buchnera sp. APS	flagellar assembly protein fliH	72	32
1017	gi15149090	lumpy skin disease virus	LSDV079 mRNA capping enzyme large subunit	66	32
1018	gi9967289	Macaca fascicularis	hypothetical protein	356	91
1019	AAG03026	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7107.	243	100
1019	gi12853136	Mus musculus	putative	166	67
1019	AAB41285	Homo sapiens	CURA- Human ORFX ORF1049 polypeptide sequence SEQ ID NO:2098.	64	34
1020	AAY36512	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 32.	748	100
1020	gi7243179	Homo sapiens	mRNA for KIAA1399 protein, partial cds.	82	41
1020	gi7243179	Homo sapiens	KIAA1399 protein	82	41
1021	AAB95621	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18338.	2058	99
1021	gi10436272	Homo sapiens	cDNA FLJ13958 fis, clone Y79AA1001216.	2058	99
1021	gi14165529	Homo sapiens	hypothetical protein FLJ12438, clone MGC:2473 IMAGE:3050071, mRNA, complete cds.	2056	99
1022	gi514268	Homo sapiens	Human proto-oncogene tyrosine-protein kinase (ABL) gene, exon 1a and exons 2-10, complete cds.	248	100
1022	gi555876	Mus musculus	c-abl protein, type IV	242	95
1022	gi49841	Mus musculus	c-abl protein	242	95
1023	AAG66758	Homo sapiens	BIOW- Human promoter binding factor 13.	627	100
1023	gi9963908	Homo sapiens	NPD009 mRNA, complete cds.	627	100
1023	gi14290450	Homo sapiens	NPD009 protein, clone MGC:16898 IMAGE:4156159, mRNA, complete cds.	624	99
1024	gi11138042	Homo sapiens	mRNA, similar to rat myomegalin, complete cds.	1227	99
1024	AAY00346	Homo sapiens	HUMA- Fragment of human secreted	1206	97

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			protein encoded by gene 2.		
1024	AAM25852	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1367.	1199	96
1025	AAG00700	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4781.	393	98
1025	gi12858787	Mus musculus	putative	313	93
1025	gi16553210	Homo sapiens	cDNA FLJ32921 fis, clone TESTI2006872.	209	70
1026	gi16924223	Homo sapiens	hypothetical protein FLJ12929, clone MGC:22200 IMAGE:4070101, mRNA, complete cds.	682	100
1026	AAB95241	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17394.	673	99
1026	gi10434702	Homo sapiens	cDNA FLJ12929 fis, clone NT2RP2004775.	673	99
1027	gi14017897	Homo sapiens	mRNA for KIAA1840 protein, partial cds.	2216	100
1027	gi10437539	Homo sapiens	cDNA: FLJ21439 fis, clone COL04352.	2210	99
1027	AAG81395	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:308.	1308	100
1028	gi3127176	Homo sapiens	sulfonylurea receptor 2B (SUR2) gene, alternatively spliced product, exon 38b and complete cds.	723	100
1028	gi3127175	Homo sapiens	sulfonylurea receptor 2A (SUR2) gene, alternatively spliced product, exon 38a and complete cds.	723	100
1028	gi15778680	Oryctolagus cuniculus	sulphonylurea receptor 2B	710	98
1029	gi14333990	Homo sapiens	enhancer of polycomb 2 (EPC2) mRNA, complete cds.	3911	99
1029	gi11907923	Homo sapiens	enhancer of polycomb mRNA, complete cds.	3879	97
1029	gi3757892	Mus musculus	enhancer of polycomb	3613	92
1030	gi9967305	Macaca fascicularis	hypothetical protein	313	94
1030	AAM80165	Homo sapiens	HYSE- Human protein SEQ ID NO 3811.	76	43
1030	AAM79181	Homo sapiens	HYSE- Human protein SEQ ID NO 1843.	76	43
1031	AAM93813	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3861.	346	95
1031	AAG01877	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5958.	341	100
1031	gi5917666	Zea mays	extensin-like protein	67	53
1032	AAM93813	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3861.	341	100
1032	AAG01877	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5958.	341	100
1032	gi10799949	Rattus norvegicus	ABC2	72	36
1033	AAY19473	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	264	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1034	gi17390437	Homo sapiens	clone MGC:9829 IMAGE:3863118, mRNA, complete cds.	879	99
1034	gi12850729	Mus musculus	putative	777	85
1034	gi10440154	Homo sapiens	cDNA: FLJ23459 fis, clone HSI07588.	758	100
1035	AAR97285	Homo sapiens	KYOW Human 26S proteasome constitutive protein P31.	1331	100
1035	gi3702282	Homo sapiens	chromosome 19, cosmid F5960, complete sequence.	1331	100
1035	gi12654653	Homo sapiens	proteasome (prosome, macropain) 26S subunit, non-ATPase, 8, clone MGC:1660 IMAGE:3528096, mRNA, complete cds.	1331	100
1036	gi12654125	Homo sapiens	hypothetical protein PP5395, clone MGC:5610 IMAGE:3461724, mRNA, complete cds.	766	99
1036	gi10441968	Homo sapiens	clone PP5395 unknown mRNA.	766	99
1036	gi12843917	Mus musculus	putative	535	73
1037	AAG02764	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6845.	281	100
1037	gi17221344	Kluyveromyces lactis	hypothetical protein	87	35
1037	gi16649041	Arabidopsis thaliana	Unknown protein	75	37
1038	AAO02417	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16309.	445	96
1038	AAG03101	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7182.	385	97
1038	gi2209200	Helobdella robusta	LOX5	73	34
1039	AAE09718	Homo sapiens	MILL- Human ubiquitin carboxy-terminal hydrolase, 23436 protein.	571	100
1039	gi16547646	Homo sapiens	unnamed protein product	571	100
1039	AAB74684	Homo sapiens	INCY- Human protease and protease inhibitor PPIM-17.	561	100
1040	AAM25866	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1381.	877	100
1040	gi10440168	Homo sapiens	cDNA: FLJ23468 fis, clone HSI11603.	877	100
1040	gi12839602	Mus musculus	putative	573	65
1041	AAB60118	Homo sapiens	INCY- Human transport protein TPPT-38.	1250	100
1041	gi16552638	Homo sapiens	cDNA FLJ32499 fis, clone SKNSH2000347, weakly similar to CYTOCHROME B2 PRECURSOR (EC 1.1.2.3).	842	98
1041	gi9801259	Leishmania major	possible CG15429 protein	449	44
1042	AAB94782	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15884.	330	70
1042	AAU27665	Homo sapiens	ZYMO Human protein AFP162878.	330	70
1042	gi15215279	Homo sapiens	hypothetical protein MGC11349, clone MGC:14984 IMAGE:3635966, mRNA, complete cds.	330	70
1043	gi10439613	Homo sapiens	cDNA: FLJ23047 fis, clone	668	99

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			LNG02513.		
1043	gi12850050	Mus musculus	putative	340	53
1043	gi13622152	Streptococcus pyogenes M1 GAS	hypothetical protein	88	29
1044	AAB94493	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15184.	193	90
1044	gi16307381	Mus musculus	Similar to dynamin 2	191	88
1044	gi12853743	Mus musculus	putative	191	88
1045	AAM25873	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1388.	516	100
1045	AAY57878	Homo sapiens	INCY- Human transmembrane protein HTMPN-2.	516	100
1045	AAU39009	Homo sapiens	GEMY Human secreted protein am728_60.	80	30
1046	AAG03414	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7495.	328	100
1046	gi2301526	unidentified	AMYLOID PROTEIN AA	100	29
1046	gi160229	Plasmodium reichenowi	circumsporozoite protein	95	30
1047	gi16550135	Homo sapiens	cDNA FLJ30851 fis, clone FEBRA2002908.	840	100
1047	gi9967240	Macaca fascicularis	hypothetical protein	557	71
1047	gi12853386	Mus musculus	putative	210	46
1048	gi3746069	Arabidopsis thaliana	putative non-LTR retroelement reverse transcriptase	74	31
1048	gi7271069	Candida albicans	hypothetical protein	71	36
1048	gi13882111	Mycobacterium tuberculosis CDC1551	PE family protein	70	34
1049	gi9947823	Pseudomonas aeruginosa	hypothetical protein	643	70
1049	gi17429445	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	365	56
1049	gi9950333	Pseudomonas aeruginosa	hypothetical protein	321	47
1050	AAY27754	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 38.	555	100
1050	gi2104464	Schizosaccharomyces pombe	hypothetical protein	71	29
1050	gi3287941	Schizosaccharomyces pombe	HYPOTHETICAL 44.3 KD PROTEIN C25H2.15 IN CHROMOSOME II >	71	29
1051	AAB95246	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17407.	757	100
1051	AAB95127	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17129.	757	100
1051	gi10434139	Homo sapiens	cDNA FLJ12572 fis, clone NT2RM4000971.	757	100
1052	AAB53066	Homo sapiens	GETH Human angiogenesis-associated protein PRO178, SEQ ID NO:11.	71	64
1052	AAB51330	Homo sapiens	HERI- Human NEW angiopoietin-like	71	64

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			protein SEQ ID NO:8.		
1052	AAY72626	Homo sapiens	HYSE- Human angiopoietin protein, CG015alt2.	71	64

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
588	BL01183	ubiE/COQ5 methyltransferase family proteins.	BL01183B 21.31 3.317e-11 146-191
629	BL00223	Annexins repeat proteins domain proteins.	BL00223A 15.59 4.414e-30 20-54 BL00223C 24.79 1.186e-11 7-62
629	PR00198	ANNEXIN TYPE II SIGNATURE	PR00198B 8.71 4.767e-13 29-52 PR00198D 7.65 4.758e-12 24-46 PR00198D 7.65 3.298e-11 96-118
629	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202B 11.44 8.986e-19 28-52 PR00202C 13.34 4.452e-16 69-86 PR00202D 5.58 5.182e-11 96-118
629	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199B 6.86 1.651e-16 29-52 PR00199D 5.65 7.039e-13 24-46 PR00199D 5.65 3.586e-10 96-118 PR00199C 13.84 7.152e-10 69-86
629	PR00197	ANNEXIN TYPE I SIGNATURE	PR00197D 7.50 8.125e-15 24-46 PR00197B 7.56 9.143e-12 29-52 PR00197D 7.50 8.813e-10 96-118
629	PR00196	ANNEXIN FAMILY SIGNATURE	PR00196A 11.16 3.700e-21 29-52 PR00196C 10.36 3.298e-17 96-118 PR00196B 10.68 7.750e-17 69-86 PR00196C 10.36 4.536e-14 24-46 PR00196E 9.19 1.563e-09 28-49
629	PR00200	ANNEXIN TYPE IV SIGNATURE	PR00200B 7.39 5.919e-15 29-52 PR00200E 10.00 5.871e-13 24-46 PR00200E 10.00 8.941e-13 96-118 PR00200D 10.01 9.471e-12 69-86 PR00200G 9.43 6.067e-09 28-55
629	PR00201	ANNEXIN TYPE V SIGNATURE	PR00201A 6.05 1.000e-28 29-52 PR00201D 10.49 3.250e-24 96-118 PR00201C 11.13 1.474e-21 69-86 PR00201B 8.88 2.552e-11 53-62 PR00201D 10.49 7.198e-09 24-46
795	BL00572	Glycosyl hydrolases family 1 proteins.	BL00572C 20.73 2.324e-25 40-75
938	PD00210	PROTEIN ANTIOXIDANT PEROXIDASE RED.	PD00210 15.25 3.912e-09 88-104
940	PD00210	PROTEIN ANTIOXIDANT PEROXIDASE RED.	PD00210 15.25 5.500e-09 88-104

* Results include in order: Accession No., subtype, e-value, and amino acid position of the signature in the corresponding polypeptide

Table 4

SEQ ID	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
527	ion_trans	Ion transport protein	8.3e-18	72.6	1	438-672
527	ank	Ankyrin repeat	1.9e-06	34.8	2	77-108:163-195
527	Srg	C.elegans Srg family integral membrane prot	8.1	-222.4	1	418-669
546	vwa	von Willebrand factor type A domain	0.77	-42.3	1	776-958
553	TPR	TPR Domain	7.1	7.4	1	40-73
566	PTPS	6-pyruvoyl tetrahydropterin synthase	0.54	-52.2	1	52-149
575	PolyA_pol	Poly A polymerase family	1.3	-61.4	1	37-155
588	Ubie_methyltran	ubiE/COQ5 methyltransferase family	0.6	-150.7	1	65-249
591	ubiquitin	Ubiquitin family	0.15	11.5	1	106-197
593	zf-C4_Topoisom	Topoisomerase DNA binding C4 zinc fing	9.2	-5.6	1	96-130
594	zf-C2H2	Zinc finger, C2H2 type	1.1	15.8	1	61-85
599	CBFD_NFYB_HMF	Histone-like transcription factor	3.8	-8.3	1	26-89
610	vwd	von Willebrand factor type D domain	7.7	-30.1	1	169-321
610	HRM	Hormone receptor domain	7.8	-13.5	1	85-150
612	Metallophos	Calcineurin-like phosphoesterase	7.9	-8.2	1	18-177
618	Peptidase_C54	Peptidase family C54	5.9e-207	700.9	1	42-332
627	AT_hook	AT hook motif	8.5	7.9	1	97-109
629	annexin	Annexin	7.6e-31	115.9	1	17-84
631	ABC-3	ABC 3 transport family	2.1	-182.9	1	152-349
631	ion_trans	Ion transport protein	8.3	-13.4	1	187-389
653	LEA	Late embryogenesis abundant protein	8.2	-6.8	1	203-270
655	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	2.9	-60.0	1	8-159
669	CBM_21	Putative phosphatase regulatory subunit	0.0056	5.1	1	280-418
671	V-ATPase_C	V-ATPase subunit C	1.3e-54	194.8	1	1-225
677	Tim17	Mitochondrial import inner membrane transloc	4e-74	259.7	1	51-184
678	Tim17	Mitochondrial import inner membrane transloc	3.1e-57	203.6	1	51-234
681	PARP	Poly(ADP-ribose) polymerase catalytic domain	5.2	-96.7	1	397-577
692	vATP-synt_E	ATP synthase (E/31 kDa) subunit	4.1	-92.4	1	276-459
693	vATP-synt_E	ATP synthase (E/31 kDa) subunit	4.1	-92.4	1	332-515
709	Ribosomal_S25	S25 ribosomal protein	7.9e-44	159.0	1	1-113
716	DUF6	Integral membrane protein DUF6	3.1	-16.3	1	11-145
717	PAP2	PAP2 superfamily	1.7	-22.6	1	174-355

Table 4

SEQ ID	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
718	IBR	IBR domain	6.2	-14.4	1	59-110
732	zf-C2H2	Zinc finger, C2H2 type	9.7	9.4	1	389-410
745	DUF81	Domain of unknown function DUF81	4.7	-44.7	1	3-150
751	Glyco_hydro_2_N	Glycosyl hydrolases family 2, sugar b	0.44	-75.0	1	37-144
761	Myc-LZ	Myc leucine zipper domain	2.2	12.8	1	136-168
762	Tropomyosin	Tropomyosin	5.2	-116.0	1	318-529
762	LEM	LEM domain	10	-4.0	1	461-504
764	SEA	SEA domain	0.076	17.1	4	112-245:270-395:561-684:955-1085
769	TTL	Tubulin-tyrosine ligase family	2.4e-93	323.5	1	35-344
780	HEAT_PBS	PBS lyase HEAT-like repeat	0.17	18.4	2	298-323:390-422
780	Adaptin_N	Adaptin N terminal region	0.46	-162.5	1	65-643
780	Dioxygenase	Dioxygenase	2.5	-106.2	1	807-937
785	CENP-B	CENP-B protein	1.4e-07	4.9	1	178-367
785	HTH_5	Bacterial regulatory protein, arsR family	0.48	5.3	1	9-93
785	HTH_3	Helix-turn-helix	1.4	10.3	1	20-74
788	zf-CCCH	Zinc finger C-x8-C-x5-C-x3-H type	1.4	9.4	2	76-103:116-144
791	Calx-beta	Calx-beta domain	0.0011	16.7	1	82-160
795	Glyco_hydro_1	Glycosyl hydrolase family 1	7.4e-07	-205.2	1	2-171
807	rrm	RNA recognition motif.	0.78	3.4	1	13-80
807	UIM	Ubiquitin interaction motif	3.4	13.2	2	650-667:673-690
822	Keratin_B2	Keratin, high sulfur B2 protein	0.15	-55.4	1	2-161
825	Acetyltransf	Acetyltransferase (GNAT) family	5.7	1.1	1	191-277
846	Glyoxalase	Glyoxalase/Bleomycin resistance protein/Di	0.074	11.7	1	2-118
850	MORN	MORN repeat	1.1e-28	108.7	7	14-36:38-59:60-80:106-128:157-179:309-331:332-354
863	NIF	NLI interacting factor	1.6e-104	360.6	1	82-256
869	Phosducin	Phosducin	0.0067	-89.2	1	1-239
870	MotA_ExbB	MotA/TolQ/ExbB proton channel family	1.5	-49.3	1	89-204
872	Armadillo_seg	Armadillo/beta-catenin-like repeat	0.42	17.1	2	677-717:727-769
872	HEAT_PBS	PBS lyase HEAT-like repeat	6.1	13.2	3	410-436:704-745:756-798
872	Adaptin_N	Adaptin N terminal region	9.5	-204.6	1	215-972
885	Patatin	Patatin-like phospholipase	9.2e-30	112.3	1	10-179
890	Glycophorin_A	Glycophorin A	5.9	-44.6	1	2-91

Table 4

SEQ ID	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
901	dehydrin	Dehydrin	6.6	-77.8	1	52-223
931	ubiquitin	Ubiquitin family	8.5	-6.5	1	405-478
938	AhpC-TSA	AhpC/TSA family	5e-05	3.6	1	58-189
940	AhpC-TSA	AhpC/TSA family	0.28	-43.3	1	58-165
943	NUDIX	MutT-like domain	8.6e-07	36.0	1	12-264
949	V-ATPase_G	Vacuolar (H+)-ATPase G subunit	5.8	-48.6	1	10-120
954	COX6C	Cytochrome c oxidase subunit VIc	1.2e-14	62.1	1	1-47
956	Myc_N_term	Myc amino-terminal region	8.1	-185.0	1	209-500
961	Cadherin_C_term	Cadherin cytoplasmic region	7.8	-83.3	1	41-148
962	cadherin	Cadherin domain	0.17	9.3	1	19-109
962	Cadherin_C_term	Cadherin cytoplasmic region	0.89	-70.9	1	159-319
963	cadherin	Cadherin domain	0.17	9.3	1	19-109
963	Cadherin_C_term	Cadherin cytoplasmic region	0.21	-62.6	1	159-301
992	Keratin_B2	Keratin, high sulfur B2 protein	5.8	-80.3	1	28-166
999	Patatin	Patatin-like phospholipase	1.4e-54	194.7	1	30-196
1001	S1	S1 RNA binding domain	1.3	4.5	1	67-131
1004	Branch	Core-2/I-Branching enzyme	0.00014	-64.7	1	3-317
1029	Mur_ligase_C	Mur ligase family, glutamate ligase doma	7.9	-11.9	1	161-235
1040	Seryl_tRNA_N	Seryl-tRNA synthetase N-terminal domain	6	0.1	1	56-102
1041	heme_1	Heme/Steroid binding domain	0.00024	22.9	1	19-98

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
527	1awc	B	18	157	8e-34	-0.23	0.52		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
527	1awc	B	3	125	4.8e-29	-0.09	0.29		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
527	1bd8		21	160	8e-27	-0.01	0.19		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
527	1bi7	B	443	560	6.4e-17	0.14	0.04		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A, MTS1; CYCLIN DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
527	1blx	B	21	160	6.4e-26	-0.17	0.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA,

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										COMPLEX (INHIBITOR PROTEIN/KINASE)
527	1blx	B	50	183	1.3e-23	0.01	0.37		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
527	1bu9	A	1	130	6.4e-25	-0.12	0.15		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN- 2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
527	1d9s	A	443	567	1.6e-17	0.14	0.06		CYCLIN-DEPENDENT KINASE 4 INHIBITOR B; CHAIN: A;	SIGNALING PROTEIN HELIX-TURN-HELIX, ANKYRIN REPEAT
527	1ihb	A	18	161	1.3e-25	-0.12	0.10		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
527	1ihb	A	1	129	3.2e-24	0.01	0.37		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
527	1myo		19	127	1.6e-20	0.14	0.01		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
543	1qcq	A	5	50	2.4e-06	0.16	0.07		UBIQUITIN	LIGASE UBIQUITIN,

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CONJUGATING ENZYME; CHAIN: A;	UBIQUITIN-CONJUGATING ENZYME, YEAST
543	2aak		1	50	1.5e-08	-0.44	0.92		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC1; UBIQUITIN CONJUGATION, LIGASE
557	1cun	A	1	217	0.0009			57.03	ALPHA SPECTRIN; CHAIN: A, B, C;	STRUCTURAL PROTEIN TWO REPEATS OF SPECTRIN, ALPHA HELICAL LINKER REGION, 2 2 TANDEM 3-HELIX COILED-COILS, STRUCTURAL PROTEIN
557	1ez3	A	37	146	9e-07	-0.01	0.49		SYNTAXIN-1A; CHAIN: A, B, C;	ENDOCYTOSIS/EXOCYTOSIS SYNAPTOTAGMIN ASSOCIATED 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE
557	1g1e	B	35	112	0.006	0.47	0.16		MAD1 PROTEIN; CHAIN: A; SIN3A; CHAIN: B;	TRANSCRIPTION MAX DIMERIZATION PROTEIN; FOUR-HELIX BUNDLE, PROTEIN-PEPTIDE COMPLEX
558	1quu	A	41	292	1.2e-06			55.25	HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
566	1b66	A	66	141	6e-07	-0.63	0.90		6-PYRUVOYL TETRAHYDROPTERIN SYNTHASE; CHAIN: A, B;	TETRAHYDROBIOPTERIN BIOSYNTHESIS TETRAHYDROBIOPTERIN BIOSYNTHESIS, PHOSPHATE ELIMINATION, 2 PTERINE SYNTHESIS
575	1cun	A	44	260	0.00015			55.09	ALPHA SPECTRIN; CHAIN: A, B, C;	STRUCTURAL PROTEIN TWO REPEATS OF SPECTRIN, ALPHA HELICAL

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										LINKER REGION, 2 2 TANDEM 3-HELIX COILED-COILS, STRUCTURAL PROTEIN
575	1quu	A	29	260	1.5e-05			64.26	HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
629	1a8a		4	117	1.5e-47			51.70	ANNEXIN V; CHAIN: NULL;	PHOSPHOLIPID ANALOG PLACENTAL ANTICOAGULANT PROTEIN; PHOSPHOLIPID ANALOG, CALCIUM BINDING PROTEIN, MEMBRANE 2 BINDING PROTEIN
629	1hvd		4	117	3.4e-49			66.12	CALCIUM/PHOSPHOLIPID BINDING ANNEXIN V (LIPOCORTIN V, ENDONEXIN II, PLACENTAL 1HVD 3 ANTICOAGULANT PROTEIN) (CALCIUM IONS ARE VISIBLE) MUTATION 1HVD 4 WITH GLU 17 REPLACED BY GLY (E17G) 1HVD 5	
630	1a8a		4	33	1.4e-09	-0.64	1.00		ANNEXIN V; CHAIN: NULL;	PHOSPHOLIPID ANALOG PLACENTAL ANTICOAGULANT PROTEIN; PHOSPHOLIPID ANALOG, CALCIUM BINDING PROTEIN, MEMBRANE 2 BINDING PROTEIN
630	1a8a		5	32	6e-10	-0.39	1.00		ANNEXIN V; CHAIN: NULL;	PHOSPHOLIPID ANALOG PLACENTAL

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										ANTICOAGULANT PROTEIN; PHOSPHOLIPID ANALOG, CALCIUM BINDING PROTEIN, MEMBRANE 2 BINDING PROTEIN
630	1hvd		4	33	3.4e-10	-0.71	0.98		CALCIUM/PHOSPHOLIPID BINDING ANNEXIN V (LIPOCORTIN V, ENDONEXIN II, PLACENTAL 1HVD 3 ANTICOAGULANT PROTEIN) (CALCIUM IONS ARE VISIBLE) MUTATION 1HVD 4 WITH GLU 17 REPLACED BY GLY (E17G) 1HVD 5	
718	1d2h	A	7	52	0.0036	-0.19	0.63		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B, C, D;	TRANSFERASE METHYLTRANSFERASE
718	1xva	A	7	52	0.0036	-0.15	0.62		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B;	METHYLTRANSFERASE GNMT, S-ADENOSYL-L-METHIONINE\: GLYCINE METHYLTRANSFERASE
737	1dfv	A	24	74	0.0032	-0.14	0.16		HUMAN NEUTROPHIL GELATINASE; CHAIN: A, B;	SUGAR BINDING PROTEIN NGAL; NEUTROPHIL, NGAL, LIPOCALIN
737	1epa	A	1	75	1.5e-10	-0.06	0.15		RETINOIC ACID-BINDING PROTEIN EPIDIDYMAL RETINOIC ACID-BINDING PROTEIN 1EPA 3 (ANDROGEN DEPENDENT SECRETORY PROTEIN) (B-FORM) 1EPA 4	
737	1qqs	A	13	75	1.8e-10	0.04	0.07		NEUTROPHIL GELATINASE; CHAIN: A;	SUGAR BINDING PROTEIN NGAL; NEUTROPHIL LIPOCALIN, SIGNAL

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										PROTEIN, GLYCOPROTEIN
737	1qqs	A	24	74	0.0019	-0.29	0.34		NEUTROPHIL GELATINASE; CHAIN: A;	SUGAR BINDING PROTEIN NGAL; NEUTROPHIL LIPOCALIN, SIGNAL PROTEIN, GLYCOPROTEIN
751	1bhg	A	72	145	6.8e-35	-0.54	0.78		BETA-GLUCURONIDASE; CHAIN: A, B;	GLYCOSIDASE GUS GENE PRODUCT; LYSOSOMAL ENZYME, ACID HYDROLASE, GLYCOSIDASE
752	1d2h	A	8	126	3.4e-17	-0.14	0.58		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B, C, D;	TRANSFERASE METHYLTRANSFERASE
752	1d2h	A	8	233	6e-22	-0.17	0.36		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B, C, D;	TRANSFERASE METHYLTRANSFERASE
752	1dus	A	15	127	1e-12	-0.08	0.89		MJ0882; CHAIN: A;	STRUCTURAL GENOMICS HYPOTHETICAL PROTEIN, METHANOCOCCUS JANNASCHII
752	1g6q	1	1	299	2.1e-82	0.57	1.00		HNRNP ARGININE N-METHYLTRANSFERASE; CHAIN: 1, 2, 3, 4, 5, 6;	TRANSFERASE SAM-BINDING DOMAIN, BETA-BARREL, MIXED ALPHA-BETA, HEXAMER, 2 DIMER
752	1g6q	1	1	301	1e-63	0.33	1.00		HNRNP ARGININE N-METHYLTRANSFERASE; CHAIN: 1, 2, 3, 4, 5, 6;	TRANSFERASE SAM-BINDING DOMAIN, BETA-BARREL, MIXED ALPHA-BETA, HEXAMER, 2 DIMER
752	1vid		25	122	9e-12	0.12	0.00		CATECHOL O-METHYLTRANSFERASE; CHAIN: NULL;	TRANSFERASE (METHYLTRANSFERASE) COMT; TRANSFERASE, METHYLTRANSFERASE, NEUROTRANSMITTER DEGRADATION
752	1xva	A	5	132	6.8e-18	-0.01	0.65		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B;	METHYLTRANSFERASE GNMT, S-ADENOSYL-L-METHIONINE, GLYCINE

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										METHYLTRANSFERASE
752	1yub		13	97	3e-07	-0.53	0.06		RRNA METHYLTRANSFERASE; CHAIN: NULL;	METHYLTRANSFERASE ERMAM; METHYLTRANSFERASE, ERM, ERMAM, MLS ANTIBIOTICS, NMR, 2 RRNA
753	1d2h	A	23	141	1.2e-16	-0.14	0.58		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B, C, D;	TRANSFERASE METHYLTRANSFERASE
753	1d2h	A	23	248	6e-22	-0.17	0.36		GLYCINE N-METHYLTRANSFERASE; CHAIN: A, B, C, D;	TRANSFERASE METHYLTRANSFERASE
753	1dus	A	28	163	1.7e-13	-0.05	0.84		MJ0882; CHAIN: A;	STRUCTURAL GENOMICS HYPOTHETICAL PROTEIN, METHANOCOCCUS JANNASCHII
753	1g6q	1	3	316	1.2e-65	0.25	1.00		HNRNP ARGININE N-METHYLTRANSFERASE; CHAIN: 1, 2, 3, 4, 5, 6;	TRANSFERASE SAM-BINDING DOMAIN, BETA-BARREL, MIXED ALPHA-BETA, HEXAMER, 2 DIMER
753	1g6q	1	4	314	2.7e-86	0.49	1.00		HNRNP ARGININE N-METHYLTRANSFERASE; CHAIN: 1, 2, 3, 4, 5, 6;	TRANSFERASE SAM-BINDING DOMAIN, BETA-BARREL, MIXED ALPHA-BETA, HEXAMER, 2 DIMER
753	1vid		40	137	9e-12	0.12	0.00		CATECHOL O-METHYLTRANSFERASE; CHAIN: NULL;	TRANSFERASE (METHYLTRANSFERASE) COMT; TRANSFERASE, METHYLTRANSFERASE, NEUROTRANSMITTER DEGRADATION
753	1vid		42	152	1.5e-05	-0.08	0.18		CATECHOL O-METHYLTRANSFERASE; CHAIN: NULL;	TRANSFERASE (METHYLTRANSFERASE) COMT; TRANSFERASE, METHYLTRANSFERASE, NEUROTRANSMITTER DEGRADATION
753	1xva	A	20	147	1.7e-17	-0.01	0.65		GLYCINE N-METHYLTRANSFERASE;	METHYLTRANSFERASE GNMT, S-ADENOSYL-L-

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CHAIN: A, B;	METHIONINE\: GLYCINE METHYLTRANSFERASE
753	1yub		30	104	6.8e-06	-0.56	0.01		RRNA METHYLTRANSFERASE; CHAIN: NULL;	METHYLTRANSFERASE ERMAM; METHYLTRANSFERASE, ERM, ERMAM, MLS ANTIBIOTICS, NMR, 2 RRNA
757	1ez3	A	68	190	9e-10	0.24	-0.12		SYNTAXIN-1A; CHAIN: A, B, C;	ENDOCYTOSIS/EXOCYTOSIS SYNAPTOTAGMIN ASSOCIATED 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE
762	1dnl	B	145	412	6e-06	-0.42	0.25		SYNTAXIN BINDING PROTEIN 1; CHAIN: A; SYNTAXIN 1A; CHAIN: B;	ENDOCYTOSIS/EXOCYTOSIS NSEC1; PROTEIN-PROTEIN COMPLEX, MULTI-SUBUNIT
762	1dvp	A	117	164	0.006	-0.71	0.15		HEPATOCYTE GROWTH FACTOR-REGULATED TYROSINE CHAIN: A;	TRANSFERASE HRS; HRS, VHS, FYVE, ZINC FINGER, SUPERHELIX
762	1f5n	A	98	263	6e-05	-0.20	0.30		INTERFERON-INDUCED GUANYLATE-BINDING PROTEIN 1; CHAIN: A;	SIGNALING PROTEIN GBP, GTP HYDROLYSIS, GDP, GMP, INTERFERON INDUCED, DYNAMIN 2 RELATED, LARGE GTPASE FAMILY. GMPPNP, GPPNHP.
773	1cun	A	83	203	0.006	-0.09	0.21		ALPHA SPECTRIN; CHAIN: A, B, C;	STRUCTURAL PROTEIN TWO REPEATS OF SPECTRIN, ALPHA HELICAL LINKER REGION, 2 2 TANDEM 3-HELIX COILED-COILS, STRUCTURAL PROTEIN
848	1a4y	A	19	482	3.4e-16			90.38	RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE),

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPITOPE MAPPING, LEUCINE-RICH 3 REPEATS
848	2bnh		1	493	3.4e-17			89.16	RIBONUCLEASE INHIBITOR; CHAIN: NULL;	ACETYLATION RNASE INHIBITOR, RIBONUCLEASE/ANGIOGEN IN INHIBITOR ACETYLATION, LEUCINE-RICH REPEATS
859	1ddm	A	43	160	3e-12	0.30	0.71		NUMB PROTEIN; CHAIN: A; NUMB ASSOCIATE KINASE; CHAIN: B;	SIGNALING PROTEIN/TRANSFERASE NAK; COMPLEX, SIGNAL TRANSDUCTION, PHOSPHOTYROSINE BINDING 2 DOMAIN (PTB), ASYMMETRIC CELL DIVISION
869	1a0r	P	20	207	6.6e-42			73.49	TRANSDUCIN; CHAIN: B, G; PHOSDUCIN; CHAIN: P;	COMPLEX (TRANSDUCER/TRANSDUCTION) GT BETA-GAMMA; MEKA, PP33 PHOSDUCIN, TRANSDUCIN, BETA-GAMMA, SIGNAL TRANSDUCTION, 2 REGULATION, PHOSPHORYLATION, G PROTEINS, THIOREDOXIN, 3 VISION, MEKA, COMPLEX (TRANSDUCER/TRANSDUCTION), 4 POST-TRANSLATIONAL MODIFICATION, FARNESYL, FARNESYLATION HEADER HETNAM
869	1b9x	C	46	207	6.8e-22			63.34	TRANSDUCIN; CHAIN: A;	SIGNALING PROTEIN GT

Table 5

SEQ ID NO	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									TRANSDUCIN; CHAIN: B; PHOSDUCIN; CHAIN: C;	BETA; GT GAMMA; MEKA, PP33; PHOSDUCIN, TRANSDUCIN, BETA-GAMMA, SIGNAL TRANSDUCTION, 2 REGULATION, PHOSPHORYLATION, G PROTEINS, THIOREDOXIN, 3 VISION, MEKA, COMPLEX (TRANSDUCER/ TRANSDUCTION), SIGNALING 4 PROTEIN
869	2trc	P	2	207	6.8e-22			63.15	TRANSDUCIN; CHAIN: B, G; PHOSDUCIN; CHAIN: P;	COMPLEX (TRANSDUCER/TRANSDUCTION) GT BETA-GAMMA; MEKA, PP33; PHOSDUCIN, TRANSDUCIN, BETA-GAMMA, SIGNAL TRANSDUCTION, 2 REGULATION, PHOSPHORYLATION, G PROTEINS, THIOREDOXIN, 3 VISION, MEKA, COMPLEX (TRANSDUCER/TRANSDUCTION)
875	1byl	A	8	200	0.0066			51.70	PIX; CHAIN: A;	TRANSPORT PROTEIN RHO-GTPASE EXCHANGE FACTOR, TRANSPORT PROTEIN

Table 6

SEQ.ID NO:	Position of Signal Peptide	Maximum score	Mean score
530	23	0.923	0.602
569	23	0.923	0.602
961	23	0.923	0.602

Table 7

SEQ ID NO:	Chromsomal location
1	10
2	13q21.1-21.3
3	21q22.3-ter
4	11
6	7q31.1-31.3
10	6
11	1
13	17
14	16p12
15	14
16	14q32
17	16p13.3
18	11cen-q12.1
19	17p13
20	13
21	6
22	9q34.13-q34.3
24	5q14
25	17
26	6p23
27	2
29	19
30	11
31	2
32	15q15-q21.1
34	13cen-13q14.2
35	21q22.3
36	21q22.3
37	X
38	8q22-q23
39	7
41	6
46	13
47	2
48	3
49	7
53	19q13.3
55	10
56	11p13
57	17
59	16
60	Xq28
61	7
62	20
63	20
64	6
65	16
66	14
67	8
68	1pter-p12
71	22
73	1p36.3-p36.2

Table 7

SEQ ID NO:	Chromsomal location
74	15q22
75	15q22
76	12
77	15
79	22q13.1
80	16
81	16
82	2q13
83	10
84	5
86	17
87	16q24.3
89	X
90	10
91	3
92	Xq22.1-22.3
94	19p13.3-p12
95	9
96	2q35
98	12p12.1
100	6p21.3
101	4q13.3
103	4q28-q32
104	4q28-q32
105	6p21.3
107	12p13
108	16
109	3
110	11
111	11
112	7q21.3-q22.1
113	11q23
115	11q22-q23
116	17
117	21q11.1
118	19
119	5
120	2
121	8p22
123	8
124	2q33
125	16
126	22q11.2
127	18p11.2
128	1q12-1q21.2
129	16
130	3q27-q28
131	5pter-p13.3
132	21q22.3
133	4q11-q13
134	17
135	xq23

Table 7

SEQ ID NO:	Chromsomal location
136	7p14-p13
137	17q21.2
140	20
142	14q24.3
143	Xp11.23
144	15
145	2
147	14q32
148	16
149	15
151	Xp11.23
152	Xp11.23
153	11
154	1p21.3-p13.1
155	5pter-p15.1
157	2p13
158	11
161	5q14.3
163	12
164	7q31.1-q31.2
166	6q25.3-27
167	6q25.3-27
169	3q26.2
170	16p13-p11
171	16p13-p11
172	16p13-p11
173	16p13-p11
175	8
176	11
178	17
180	15
181	15
182	16p13.3-p12
183	11q23.3
184	7
185	20q12-13.1
186	12
188	18
189	6p21.3
190	3
191	20
192	8
193	6p21.3
196	19p13.1
197	1p12
198	Xq21.1-Xq21.3
200	4
201	19
204	3
205	3p13-q26.1
206	14q21.1-q24.2
208	Y

Table 7

SEQ ID NO:	Chromsomal location
209	6q16.1-q16.3
210	16
213	17
214	18q12
215	11p11-q11
216	11p11-q11
217	7
218	3p13-q26.1
219	19q13.2
222	2
225	22q11.23
226	19
227	19
229	6
230	1p35.1-36.11
231	17
233	17q11
235	Xq27.3-28
236	19q13.3-q13.4
237	X
238	19
239	7q11-q22
240	5
241	7q21.2-q31.1
242	7
243	20q11.2
244	2p23.3-q24.3
245	17
246	3
247	1q12
248	17
249	22
253	9
254	11q
255	17
256	6p12.1-21.1
257	20
258	7
259	11q13
262	16
264	3p
265	4
266	9
267	1
268	7
272	6
273	2q37
274	19
276	19
277	17
278	10pter-q22.1
282	14

Table 7

SEQ ID NO:	Chromsomal location
283	11
285	13
286	1
287	2p12
288	17
292	5
293	10q25.1
294	15
295	17
296	21q22.1
297	4
300	1p36.11-36.2
302	9q13-q21
303	5
304	4q25
305	17
306	5
307	7
308	2
309	10
311	14
312	11q23
313	17
314	17
316	5p15.2-q12.3
317	10
319	11q24
321	11
322	17
323	1
324	8q21
325	6p21.2-21.3
326	11
327	1p
328	Xq21.1-21.32
329	10
331	14q24.3
332	11q24
333	17
336	5
337	2q35
338	11
339	19
340	11
341	22
342	14q21.1-q21.3
343	8q23
344	22q13.1
345	22q11.23
346	17
347	18
348	5

Table 7

SEQ ID NO:	Chromsomal location
349	17
350	2
351	20
353	1q12
355	18p11.2
356	10
357	15
359	22q13.1-13.32
361	1q21
363	6
364	1
365	3q23-q25
366	6p21.1
367	11p15
368	19q13
369	17
370	18q23
371	12
373	1
374	17
375	12q
376	10cen-q26.11
377	17p13.3
379	13
380	22q11.2
381	17p11.2
382	11q13
383	17
384	2p11
387	15
388	7p14-p13
389	13cen-13q14.2
390	11
391	11
393	Xq21.1-21.33
394	4
396	17
397	2p22-p21
398	2
399	17
400	19
401	15
404	17
405	8q13
406	q34.11-34.3
408	17
410	8q22-q23
412	11q13
413	11q13
414	11q13
415	16q24.3
416	12q13.1

Table 7

SEQ ID NO:	Chromsomal location
417	19
418	15
419	17
422	21q22.3
425	8
426	9
427	2q36
430	13q14.2-21.1
431	13q12.2-q13.3
433	15
434	6q26-q27
435	20q13.2-q13.33
436	20q13.2-q13.33
437	20q13.2-q13.33
438	15
440	8q22.3
442	3q29-qter
443	1
444	16
445	16
446	2p23.3-q34
447	2p23.3-q34
448	16
449	17
452	14
453	22q12
454	15
455	21q22.3
456	20
457	17p11.2
460	17
461	14q21
463	20p11.2
465	11q13
466	17
467	20
468	11q13
469	17
473	17
474	8q22-q23
475	8q21.1-q21.2
476	7q21
477	17
478	X
480	11q12.2
482	6
483	12p13
484	5
485	13
487	Xp11.23
489	1q32.1
490	Xq21.1

Table 7

SEQ ID NO:	Chromsomal location
494	13
495	17
496	9q34.1
497	19p13.3
498	1q21.2-q22
500	17
501	15
502	12p12.1
503	10p11
504	2
508	2
509	19p11-q11
510	10
511	19
512	10
514	4
515	17
516	3p13-q26.1
517	4
518	19
520	17
525	15

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
1	527	1
2	528	2
3	529	3
4	530	4
5	531	5
6	532	6
7	533	7
8	534	8
9	535	9
10	536	10
11	537	11
12	538	12
13	539	13
14	540	14
15	541	15
16	542	16
17	543	17
18	544	18
19	545	19
20	546	20
21	547	21
22	548	22
23	549	23
24	550	24
25	551	25
26	552	26
27	553	27
28	554	28
29	555	29
30	556	30
31	557	31
32	558	32
33	559	33
34	560	34
35	561	35
36	562	36
37	563	37
38	564	38
39	565	39
40	566	40
41	567	41
42	568	42
43	569	43
44	570	44
45	571	45
46	572	46
47	573	47
48	574	48
49	575	49
50	576	50
51	577	51
52	578	52

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
53	579	53
54	580	54
55	581	55
56	582	56
57	583	57
58	584	58
59	585	59
60	586	60
61	587	61
62	588	62
63	589	63
64	590	64
65	591	65
66	592	66
67	593	67
68	594	68
69	595	69
70	596	70
71	597	71
72	598	72
73	599	73
74	600	74
75	601	75
76	602	76
77	603	77
78	604	78
79	605	79
80	606	80
81	607	81
82	608	82
83	609	83
84	610	84
85	611	85
86	612	86
87	613	87
88	614	88
89	615	89
90	616	90
91	617	91
92	618	92
93	619	93
94	620	94
95	621	95
96	622	96
97	623	97
98	624	98
99	625	99
100	626	100
101	627	101
102	628	102
103	629	103
104	630	104

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
105	631	105
106	632	106
107	633	107
108	634	108
109	635	109
110	636	110
111	637	111
112	638	112
113	639	113
114	640	114
115	641	115
116	642	116
117	643	117
118	644	118
119	645	119
120	646	120
121	647	121
122	648	122
123	649	123
124	650	124
125	651	125
126	652	126
127	653	127
128	654	128
129	655	129
130	656	130
131	657	131
132	658	132
133	659	133
134	660	134
135	661	135
136	662	136
137	663	137
138	664	138
139	665	139
140	666	140
141	667	141
142	668	142
143	669	143
144	670	144
145	671	145
146	672	146
147	673	147
148	674	148
149	675	149
150	676	150
151	677	151
152	678	152
153	679	153
154	680	154
155	681	155
156	682	156

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
157	683	157
158	684	158
159	685	159
160	686	160
161	687	161
162	688	162
163	689	163
164	690	164
165	691	165
166	692	166
167	693	167
168	694	168
169	695	169
170	696	170
171	697	171
172	698	172
173	699	173
174	700	174
175	701	175
176	702	176
177	703	177
178	704	178
179	705	179
180	706	180
181	707	181
182	708	182
183	709	183
184	710	184
185	711	185
186	712	186
187	713	187
188	714	188
189	715	189
190	716	190
191	717	191
192	718	192
193	719	193
194	720	194
195	721	195
196	722	196
197	723	197
198	724	198
199	725	199
200	726	200
201	727	201
202	728	202
203	729	203
204	730	204
205	731	205
206	732	206
207	733	207
208	734	208

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
209	735	209
210	736	210
211	737	211
212	738	212
213	739	213
214	740	214
215	741	215
216	742	216
217	743	217
218	744	218
219	745	219
220	746	220
221	747	221
222	748	222
223	749	223
224	750	224
225	751	225
226	752	226
227	753	227
228	754	228
229	755	229
230	756	230
231	757	231
232	758	232
233	759	233
234	760	234
235	761	235
236	762	236
237	763	237
238	764	238
239	765	239
240	766	240
241	767	241
242	768	242
243	769	243
244	770	244
245	771	245
246	772	246
247	773	247
248	774	248
249	775	249
250	776	250
251	777	251
252	778	252
253	779	253
254	780	254
255	781	255
256	782	256
257	783	257
258	784	258
259	785	259
260	786	260

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
261	787	261
262	788	262
263	789	263
264	790	264
265	791	265
266	792	266
267	793	267
268	794	268
269	795	269
270	796	270
271	797	271
272	798	272
273	799	273
274	800	274
275	801	275
276	802	276
277	803	277
278	804	278
279	805	279
280	806	280
281	807	281
282	808	282
283	809	283
284	810	284
285	811	285
286	812	286
287	813	287
288	814	288
289	815	289
290	816	290
291	817	291
292	818	292
293	819	293
294	820	294
295	821	295
296	822	296
297	823	297
298	824	298
299	825	299
300	826	300
301	827	301
302	828	302
303	829	303
304	830	304
305	831	305
306	832	306
307	833	307
308	834	308
309	835	309
310	836	310
311	837	311
312	838	312

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
313	839	313
314	840	314
315	841	315
316	842	316
317	843	317
318	844	318
319	845	319
320	846	320
321	847	321
322	848	322
323	849	323
324	850	324
325	851	325
326	852	326
327	853	327
328	854	328
329	855	329
330	856	330
331	857	331
332	858	332
333	859	333
334	860	334
335	861	335
336	862	336
337	863	337
338	864	338
339	865	339
340	866	340
341	867	341
342	868	342
343	869	343
344	870	344
345	871	345
346	872	346
347	873	347
348	874	348
349	875	349
350	876	350
351	877	351
352	878	352
353	879	353
354	880	354
355	881	355
356	882	356
357	883	357
358	884	358
359	885	359
360	886	360
361	887	361
362	888	362
363	889	363
364	890	364

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
365	891	365
366	892	366
367	893	367
368	894	368
369	895	369
370	896	370
371	897	371
372	898	372
373	899	373
374	900	374
375	901	375
376	902	376
377	903	377
378	904	378
379	905	379
380	906	380
381	907	381
382	908	382
383	909	383
384	910	384
385	911	385
386	912	386
387	913	387
388	914	388
389	915	389
390	916	390
391	917	391
392	918	392
393	919	393
394	920	394
395	921	395
396	922	396
397	923	397
398	924	398
399	925	399
400	926	400
401	927	401
402	928	402
403	929	403
404	930	404
405	931	405
406	932	406
407	933	407
408	934	408
409	935	409
410	936	410
411	937	411
412	938	412
413	939	413
414	940	414
415	941	415
416	942	416

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
417	943	417
418	944	418
419	945	419
420	946	420
421	947	421
422	948	422
423	949	423
424	950	424
425	951	425
426	952	426
427	953	427
428	954	428
429	955	429
430	956	430
431	957	431
432	958	432
433	959	433
434	960	434
435	961	435
436	962	436
437	963	437
438	964	438
439	965	439
440	966	440
441	967	441
442	968	442
443	969	443
444	970	444
445	971	445
446	972	446
447	973	447
448	974	448
449	975	449
450	976	450
451	977	451
452	978	452
453	979	453
454	980	454
455	981	455
456	982	456
457	983	457
458	984	458
459	985	459
460	986	460
461	987	461
462	988	462
463	989	463
464	990	464
465	991	465
466	992	466
467	993	467
468	994	468

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
469	995	469
470	996	470
471	997	471
472	998	472
473	999	473
474	1000	474
475	1001	475
476	1002	476
477	1003	477
478	1004	478
479	1005	479
480	1006	480
481	1007	481
482	1008	482
483	1009	483
484	1010	484
485	1011	485
486	1012	486
487	1013	487
488	1014	488
489	1015	489
490	1016	490
491	1017	491
492	1018	492
493	1019	493
494	1020	494
495	1021	495
496	1022	496
497	1023	497
498	1024	498
499	1025	499
500	1026	500
501	1027	501
502	1028	502
503	1029	503
504	1030	504
505	1031	505
506	1032	506
507	1033	507
508	1034	508
509	1035	509
510	1036	510
511	1037	511
512	1038	512
513	1039	513
514	1040	514
515	1041	515
516	1042	516
517	1043	517
518	1044	518
519	1045	519
520	1046	520

Table 8

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 09/810,173
521	1047	521
522	1048	522
523	1049	523
524	1050	524
525	1051	525
526	1052	526

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 – 526, a mature protein coding portion of SEQ ID NO: 1 – 526, an active domain coding protein of SEQ ID NO: 1 – 526, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

3. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

4. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

5. A vector comprising the polynucleotide of claim 1.

6. An expression vector comprising the polynucleotide of claim 1.

7. A host cell genetically engineered to comprise the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

9. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of a polypeptide encoded by any one of the polynucleotides of claim 1 i.e. SEQ ID NO: 527 - 1052).

10. A composition comprising the polypeptide of claim 9 and a carrier.

11. An antibody directed against the polypeptide of claim 9.

12. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

14. The method of claim 13, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

15. A method for detecting the polypeptide of claim 9 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 9 is detected.

16. A method for identifying a compound that binds to the polypeptide of claim 9, comprising:

a) contacting the compound with the polypeptide of claim 9 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 9 is identified.

17. A method for identifying a compound that binds to the polypeptide of claim 9, comprising:

a) contacting the compound with the polypeptide of claim 9, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 9 is identified.

18. A method of producing the polypeptide of claim 9, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-526, a mature protein coding portion of SEQ ID NO: 1-526, an active domain coding portion of SEQ ID NO: 1-526, complementary sequences thereof, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

19. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides from the Sequence Listing, the mature protein portion thereof, or the active domain thereof.

20. The polypeptide of claim 21 wherein the polypeptide is provided on a polypeptide array.

21. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO: 1 – 526.

22. The collection of claim 21, wherein the collection is provided on a nucleic acid array.

23. The collection of claim 22, wherein the array detects full-matches to any one of the polynucleotides in the collection.

24. The collection of claim 22, wherein the array detects mismatches to any one of the polynucleotides in the collection.

25. The collection of claim 21, wherein the collection is provided in a computer-readable format.

26. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 9 or 19 and a pharmaceutically acceptable carrier.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 9 or 19 and a pharmaceutically acceptable carrier.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/05109

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/12, 15/00, 1/21; C12P 21/02; G01N 33/48

US CL : 536/23.1, 23.5; 435/320.1, 252.3, 69.1; 702/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.5; 435/320.1, 252.3, 69.1; 702/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Compugen (nucleic acid and amino acid sequence databases) SEQ ID NO: 1

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Databas Gencore on EST, AN AJ272034, MURPHY et al. 'Distribution of a novel human capacitative calcium entry channel, htrp7', Gene Sequence, 11 August 2000.	2

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

17 July 2002 (17.07.2002)

Date of mailing of the international search report

09 AUG 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

James Martinell Felicia D. Roberts for

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/05109

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-8, 12-14, 18, and 21-25 insofar as they relate to SEQ ID NO: 1.

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/05109

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-8, 12-14, 18, and 21-25, drawn to polynucleotides, vectors, host cells, methods of producing polypeptides, and methods of detecting polynucleotides.

Group II, claim(s) 9-11, 15-17, 19, 20, 26, and 27, drawn to polypeptides, polypeptide arrays, methods of detecting polypeptides, methods of identifying compounds that bind to polypeptides, methods of treatment using polypeptides, antibodies, and methods of using antibodies.

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

The polynucleotides, vectors, host cells, methods of producing polypeptides, and methods of detecting polynucleotides of Group I are not needed to produce or use the polypeptides or antibodies of Group II, which polypeptides may be isolated from naturally occurring sources or may be synthesized chemically.

Each of the two major groups above requires either 526 unrelated, independent, and distinct (from one another) polynucleotides (Group I) or 526 unrelated, independent, and distinct (from one another) polypeptides (Group II). For either of the two groups to be examined, applicant is further required to elect for search, one polynucleotide sequence or one polypeptide sequence, as appropraiate, within the group. See MPEP 1850.